

Министерство науки и высшего образования Российской Федерации
федеральное государственное автономное образовательное учреждение
высшего образования
**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»**

Школа инженерного предпринимательства

Направление подготовки 27.04.04 Управление в технических системах

МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

Тема работы
Разработка информационной системы управления эффективностью диверсификации экономики моногорода

УДК: 004.658:338.33:334.724.2

Студент

Группа	ФИО	Подпись	Дата
ЗВМ81	Малеева Е.А.		

Руководитель

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент	Антонова И.С.	к.э.н., доцент		

КОНСУЛЬТАНТЫ:

По разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент	Кашук И.В.	к.т.н.		

По разделу «Социальная ответственность»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент	Белоенко Е.В.	к.т.н.		

ДОПУСТИТЬ К ЗАЩИТЕ:

Руководитель ООП	ФИО	Ученая степень, звание	Подпись	Дата
Доцент	Жданова А.Б.	к.э.н.		

**Планируемые результаты обучения по направлению
27.04.04 Управление в технических системах**

Код	Результат обучения
Общие по направлению подготовки	
P1	Применять глубокие естественнонаучные и математические знания для решения научных и инженерных задач в области анализа, синтеза, проектирования, производства и эксплуатации средств автоматизации и систем управления техническими объектами
P2	Обрабатывать, анализировать и обобщать научно-техническую информацию, передовой отечественный и зарубежный опыт в области теории, проектирования, производства и эксплуатации средств автоматизации и систем управления техническими объектами
P3	Выполнять инновационные инженерные проекты по разработке программно-аппаратных средств автоматизированных систем различного назначения с использованием аналитических методов, сложных моделей, современных методов проектирования, систем автоматизированного проектирования и передового опыта разработки конкурентоспособных изделий
P4	Планировать и проводить теоретические и экспериментальные исследования в области проектирования аппаратных и программных средств автоматизированных систем с использованием новейших достижений науки и техники, передового отечественного и зарубежного опыта. Критически оценивать полученные данные и делать выводы
P5	Эффективно работать индивидуально и в качестве члена (руководителя) профессиональной междисциплинарной и международной группы; владеть иностранным языком на уровне, позволяющем работать в интернациональной профессиональной среде с пониманием культурных, языковых и социально-экономических различий
P6	Иметь широкую эрудицию, в том числе знание и понимание современных общественных и политических проблем, вопросов безопасности и охраны здоровья сотрудников, юридических аспектов, ответственности за инженерную деятельность, влияния инженерных решений на социальный контекст и окружающую среду
P7	Применять навыки управления разработкой и производством продукции на всех этапах ее жизненного цикла с учетом инновационных рисков коммерциализации проектов, в том числе в условиях неопределенности
P8	Демонстрировать способность к самостоятельному обучению, непрерывному самосовершенствованию в инженерной деятельности.
Профиль «Прикладной системный инжиниринг»	
P11	Иметь навыки управления проектами по разработке и внедрению систем автоматического и автоматизированного управления, уметь планировать этапы и мероприятия в рамках выполнения проекта, обеспечивать взаимодействие между участниками проекта, планировать потребность в ресурсах, составлять бюджет проекта, оценивать риски и их влияние на реализацию проекта.
P12	Иметь навыки формализации бизнес-процессов промышленного предприятия, уметь определять параметры бизнес-процессов, уметь применять ERP-систем для контроля бизнес-процессом, уметь проводить оценку и оптимизацию бизнес-процессов во взаимосвязи с целями предприятия и устанавливать KPI руководителям процессов.

Министерство науки и высшего образования Российской Федерации

федеральное государственное автономное образовательное учреждение

высшего образования

**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»**

Школа инженерного предпринимательства

Направление подготовки 27.04.04 Управление в технических системах

УТВЕРЖДАЮ:

Руководитель ООП

_____ Жданова А.Б.

(Подпись) (Дата) (Ф.И.О.)

ЗАДАНИЕ

на выполнение выпускной квалификационной работы

В форме:

Магистерской диссертации

(бакалаврской работы/магистерской диссертации)

Студенту:

Группа	ФИО
ЗВМ81	Малеевой Екатерине Александровне

Тема работы:

Разработка информационной системы управления эффективностью диверсификации экономики моногорода	
Утверждена приказом директора (дата, номер)	08.05.2020, № 129-9/с

Срок сдачи студентом выполненной работы:	
--	--

ТЕХНИЧЕСКОЕ ЗАДАНИЕ

Исходные данные к работе (наименование объекта исследования или проектирования; производительность или нагрузка; режим работы (непрерывный, периодический, циклический и т. д.); вид сырья или материал изделия; требования к продукту, изделию или	Объектом исследования являются монопрофильные муниципальные образования (моногорода) Российской Федерации на примере Тайгинского городского округа.
---	---

<p><i>процессу; особые требования к особенностям функционирования (эксплуатации) объекта или изделия в плане безопасности эксплуатации, влияния на окружающую среду, энергозатратам; экономический анализ и т. д.).</i></p>	
<p>Перечень подлежащих исследованию, проектированию и разработке вопросов <i>(аналитический обзор по литературным источникам с целью выяснения достижений мировой науки техники в рассматриваемой области; постановка задачи исследования, проектирования, конструирования; содержание процедуры исследования, проектирования, конструирования; обсуждение результатов выполненной работы; наименование дополнительных разделов, подлежащих разработке; заключение по работе).</i></p>	<p>Понятие и критерии выделения моногородов, получить и протестировать модели машинного обучения для предсказания доли моноотрасли Тайгинского городского округа.</p>
<p>Перечень графического материала <i>(с точным указанием обязательных чертежей)</i></p>	<p>Рисунок 1 – Разделение количества моногородов по категориям Рисунок 2 – Распределение моногородов по регионам Рисунок 3 – Эффективность диверсификации монопрофильных городов по энтропии и динамике совокупной выручки за 2013-2017 гг. Рисунок 4 - Эффективность диверсификации монопрофильных городов по энтропии и динамике деловой активности за 2013-2017 гг Рисунок 5 – Динамика численности населения Рисунок 6 - Диаграмма размаха выручки предприятий Тайгинского городского округа Рисунок 7 - Описательная статистика разброса выручки предприятий Тайгинского городского округа Рисунок 8 - Диаграмма размаха основных средств предприятий Тайгинского городского округа Рисунок 9 - Описательная статистика разброса основных средств предприятий Тайгинского городского округа Рисунок 10 - Диаграмма размаха оплаты труда предприятий Тайгинского городского округа Рисунок 11 - Описательная статистика разброса основных средств предприятий Тайгинского городского округа Рисунок 12 – Учебный набор данных подержанных автомобилей и соответствующая функция. Для простоты, Рисунок 13 – Схема нейронной сети прямого распространения Рисунок 14 – Место экономических информационных систем в контуре систем управления Рисунок 15 – Фрагмент матрицы корреляций Рисунок 16 – Программный код разделения выборки на обучающую и тестовую Рисунок 17 – Программный код для построения моделей для доли моноотрасли по основным средствам Рисунок 18 - Программный код для построения моделей для доли моноотрасли по выручке</p>

	Рисунок 19 - Программный код для построения моделей для доли моноотрасли по оплате труда
Консультанты по разделам выпускной квалификационной работы (с указанием разделов)	
Раздел	Консультант
Финансовый менеджмент, ресурсоэффективность и ресурсосбережение	Кащук И.В.
Социальная ответственность	Белоевко Е.В.
Названия разделов, которые должны быть написаны на русском и иностранном языках:	
Моделирование процесса оценки уровня диверсификации моногорода на примере Тайгинского городского округа	Modeling of the process of assessing the level of diversification of a single-industry town using the example of Taiginsky district
Дата выдачи задания на выполнение выпускной квалификационной работы по линейному графику	

Задание выдал руководитель:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент	Антонова И.С.	к.э.н., доцент		

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
ЗВМ81	Малеева Е.А.		

Министерство науки и высшего образования Российской Федерации
федеральное государственное автономное образовательное учреждение
высшего образования
**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»**

Школа инженерного предпринимательства
Направление подготовки 27.04.04 Управление в технических системах
Уровень образования магистратура
Период выполнения (осенний / весенний семестр 2019/2020 учебного года)

Форма представления работы:

Магистерская диссертация

Разработка информационной системы управления эффективностью диверсификации экономики моногорода
--

КАЛЕНДАРНЫЙ РЕЙТИНГ-ПЛАН
выполнения выпускной квалификационной работы

Срок сдачи студентом выполненной работы:	
--	--

Дата контроля	Название раздела (модуля) / вид работы (исследования)	Максимальный балл раздела (модуля)
20.03.20	Раздел 1. Проблемы развития моногородов на основе диверсификации экономики	20
17.04.20	Раздел 2. Моделирование процесса оценки уровня диверсификации моногорода на примере Тайгинского городского округа	30
11.05.20	Раздел 3. Разработка модели машинного обучения для предсказания значения доли моноотрасли на примере Тайгинского городского округа	30
25.05.20	Финансовый менеджмент, ресурсоэффективность и ресурсосбережение.	10
25.05.20	Социальная ответственность	10

Составил преподаватель:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент	Антонова И.С.	к.э.н., доцент		

Принял студент:

ФИО	Подпись	Дата
Малеева Е.А.		

СОГЛАСОВАНО:

Руководитель ООП	ФИО	Ученая степень, звание	Подпись	Дата
Доцент	Жданова А.Б.	к.э.н.		

**ЗАДАНИЕ ДЛЯ РАЗДЕЛА
«ФИНАНСОВЫЙ МЕНЕДЖМЕНТ, РЕСУРСОЭФФЕКТИВНОСТЬ И
РЕСУРСОСБЕРЕЖЕНИЕ»**

Студенту:

Группа	ФИО
ЗВМ81	Малеевой Екатерине Александровне

Школа	ШИП	Отделение школы (НОЦ)	
Уровень образования	Магистратура	Направление/специальность	27.04.04 Управление технических системах в

Исходные данные к разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»:

1. Стоимость ресурсов научного исследования (НИ): материально-технических, энергетических, финансовых, информационных и человеческих	Стоимость материальных ресурсов и специального оборудования определены в соответствии с рыночными ценами г. Томска. Тарифные ставки исполнителей определены штатным расписанием НИ ТПУ.
2. Нормы и нормативы расходования ресурсов	Тариф на расходы электроэнергии, коэффициенты для расчета заработной платы
3. Используемая система налогообложения, ставки налогов, отчислений, дисконтирования и кредитования	Отчисления во внебюджетные фонды 30 %

Перечень вопросов, подлежащих исследованию, проектированию и разработке:

1. Анализ конкурентных технических решений (НИ)	Расчет конкурентоспособности SWOT-анализ
2. Формирование плана и графика разработки и внедрения (НИ)	Структура работ. Определение трудоемкости. Разработка графика проведения исследования.
3. Составление бюджета (НИ)	Расчет бюджетной стоимости НИ
4. Оценка ресурсной, финансовой, бюджетной эффективности (НИ)	Интегральный финансовый показатель. Интегральный показатель ресурсоэффективности. Интегральный показатель эффективности.

Перечень графического материала (с точным указанием обязательных чертежей)

1. Оценка конкурентоспособности НИ
2. Матрица SWOT
3. Диаграмма Ганта
4. Бюджет НИ
5. Основные показатели эффективности НИ

Дата выдачи задания для раздела по линейному графику

Задание выдал консультант:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент	Кащук Ирина Вадимовна	К.Т.Н		

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
ЗВМ81	Малеева Екатерина Александровна		

ЗАДАНИЕ ДЛЯ РАЗДЕЛА «СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ»

Студенту:

Группа	ФИО
3ВМ81	Малеевой Екатерине Александровне

Школа	ШИП	Отделение (НОЦ)	
Уровень образования	Магистратура	Направление/специальность	27.04.04 «Управление в технических системах»

Тема ВКР:

Разработка информационной системы управления эффективностью диверсификации экономики моногорода

Исходные данные к разделу «Социальная ответственность»:

1. Характеристика объекта исследования (вещество, материал, прибор, алгоритм, рабочая зона) и области его применения	- оценка эффективности диверсификации моногорода; - разработка может быть реализована в процессе стратегического планирования администрациями моногородов, в научно-исследовательской деятельности.
--	--

Перечень вопросов, подлежащих исследованию, проектированию и разработке:

1. Правовые и организационные вопросы обеспечения безопасности: <ul style="list-style-type: none"> – специальные (характерные при эксплуатации объекта исследования, проектируемой рабочей зоны) правовые нормы трудового законодательства; – организационные мероприятия при компоновке рабочей зоны. 	<ul style="list-style-type: none"> - Трудовой кодекс Российской Федерации от 30.12.2001 197-ФЗ (ред. от 16.12.2019); - СанПиН 2.2.2/2.4.1340-03; - ГОСТ 12.2.032-78. - СанПиН 2.2.4.548-96; - ГОСТ 12.1.003-2014 - СанПиН 2.2.4.3359-16; - ГОСТ 12.1.006-84.
2. Производственная безопасность: 2.1. Анализ выявленных вредных и опасных факторов 2.2. Обоснование мероприятий по снижению воздействия	<ul style="list-style-type: none"> – Отклонение показателей микроклимата. – Превышение уровня шума. – Недостаточная освещенность рабочей зоны. – Повышенный уровень электромагнитных излучений. – Повышенное значение напряжения в электрической цепи, замыкание которой может произойти через тело человека
3. Экологическая безопасность:	Анализ воздействия на литосферу.
4. Безопасность в чрезвычайных ситуациях:	Выбор и описание возможных ЧС; типичная ЧС – пожар. - разработка превентивных мер по предупреждению ЧС; - разработка действий в результате возникшей ЧС и мер по ликвидации её последствий.

Дата выдачи задания для раздела по линейному графику	
--	--

Задание выдал консультант:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент	Белоевко Е.В.	К.Т.Н.		

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
3ВМ81	Малеева Е.А.		

Реферат

Диссертация содержит 95 страниц, 20 рисунков, 29 таблиц, 54 использованных источников, 2 приложения.

Ключевые слова: моногород, Тайгинский городского округ, диверсификация, машинное обучение, язык программирования Python.

Объектом исследования являются монопрофильные муниципальные образования (моногорода) Российской Федерации, в т.ч. Тайгинский городской округ.

Цель работы: разработать элементы информационной системы управления эффективностью диверсификации моногорода при отсутствии данных о моноотрасли, включая процесс обработки данных финансовой отчетности предприятий с применением методов машинного обучения и регрессионную модель оценки уровня диверсификации экономики моногорода.

В результате исследования была проведена обработка данных финансовой отчетности предприятий из системы СПАРК и разработана математическая модель, позволяющая оценивать уровень диверсификации моногорода при отсутствии данных о моноотрасли (в случае, если градообразующее предприятие является филиалом другой организации или структурной, расположенной вне муниципального образования).

Область применения: научно-исследовательская деятельность, в процессе стратегического планирования администрации моногородов.

Значимость работы состоит в упрощении проведения расчетов показателей концентрации и диверсификации для моногородов, заполнение нулевых значений доли моноотрасли.

В будущем планируется создание экономической информационной системы для анализа данных и расчета показателей для моногородов.

Оглавление

Реферат	11
Введение	13
1 Проблемы развития моногородов на основе диверсификации экономики .	16
1.1 Понятие, критерии выделения и классификация моногородов.....	16
1.2 Методические основы оценки эффективности диверсификации	19
1.3 Сравнительный анализ эффективности диверсификации моногородов РФ.....	23
2 Моделирование процесса оценки уровня диверсификации моногорода на примере Тайгинского городского округа	31
2.1 Общая характеристика Тайгинского городского округа	31
2.2 Систематизация методов машинного обучения для поддержки принятия решения об уровне и направлениях диверсификации моногорода	36
2.3 Обоснование языка программирования для разработки модели машинного обучения	42
3.1 Экономические информационные системы как элемент поддержки принятия управленческого решения	44
3.2 Модель машинного обучения как элемент системы поддержки принятия управленческого решения	47
3.3 Система поддержки оценки и принятия решения о диверсификации моногорода.....	49
4 Финансовый менеджмент, ресурсоэффективность и ресурсосбережение...	63
5 Социальная ответственность	77
Заключение	88
Список публикаций студента.....	89
Список литературы	90
Приложение А	96
Приложение Б	97

Введение

Проблема развития моногородов России актуальна, так как в 323 моногородах проживает более 13,5 млн человек, или 9,2% населения страны. А большинство из этих городов находятся в сложном социально-экономическом положении. Правительство в распоряжении от 2014 года отнесло к наиболее кризисным 75 городов, на данный момент из число увеличилось на четверть.

Поддержка экономики моногородов является одной из главных задач для государства. Для этого была разработана госпрограмма на 2019-2024 гг., создан Фонд социального развития моногородов, проведена подготовка большинства управленческих команд, разработаны Комплексные инвестиционные планы модернизации моногородов, создаются территории опережающего социально-экономического развития (ТОСЭР), осуществляется софинансирование инфраструктурных проектов из Фонда развития моногородов, разработана программа кредитования проектов диверсификации моногородов, профинансированы отдельные инвестиционные проекты. Однако текущая программа развития, а также комплекс мероприятий носят не опережающий, а догоняющий характер, учитывая практику реструктуризации регионов ресурсного типа в Канаде, США, Австралии, Германии, Англии и прочих Европейских странах середины прошлого века.

Актуальностью данной работы является то, что не смотря на совокупность принимаемых мер по поддержке и развитию моногородов, тем не менее вопросы качественной оценки и принятия решений остаются на низком уровне. Так, в ряде моногородов наблюдается недостаток данных по ключевому показателю – доле моноотрасли, что не позволяет проводить качественный сопоставительный анализ и оценивать уровень эффективности диверсификации. Напротив, основные показатели эффективности реализации программы развития моногородов – число вновь созданных мест и уровень безработицы – показывают невозможность проверки полученных

результатов (о чем заявляют ФНС и Счетная палата) [27]. Отсутствие единого представления об оценке и анализе данного процесса не позволяет сопоставить процессы трансформации, протекающие в моногородах. Разрозненность и фрагментарность данных о социально-экономическом развитии моногородов снижает качество проводимых оценок. В таких условиях разработка информационной системы обработки и анализа данных на микроуровне позволит оценить процессы диверсификации в моногородах и выявить закономерности их развития, предложить управленческие методы, направленные на совершенствование данного процесса.

Цель работы: разработать элементы информационной системы управления эффективностью диверсификации моногорода при отсутствии данных о моноотрасли, включая процесс обработки данных финансовой отчетности предприятий с применением методов машинного обучения и регрессионную модель оценки уровня диверсификации экономики моногорода.

Для достижения поставленной цели были сформулированы следующие задачи:

1. Определить понятие и критерии выделения моногородов.
2. Рассмотреть методы оценки эффективности диверсификации.
3. Проанализировать моногорода Кемеровской, Челябинской и Свердловской областей на оценку эффективности диверсификации.
4. Дать общую характеристику Тайгинскому городскому округу.
5. Рассмотреть алгоритмы метода машинного обучения.
6. Обосновать выбор языка программирования.
7. Разработать программу для автоматического расчета индексов концентрации для моногородов.
8. Получить и протестировать модели машинного обучения по рассчитанным данным для предсказания доли моноотрасли Тайгинского городского округа.

Объектом исследования являются монопрофильные муниципальные образования (моногорода) Российской Федерации в т.ч. Тайгинский городской округа.

Предметом исследования является процесс оценки уровня диверсификации экономики моногорода.

Научная новизна заключается в выявлении факторов и разработке математических инструментов оценки уровня диверсификации экономики моногорода для преодоления барьера отсутствия данных о моноотрасли (в случае, если градообразующее предприятие является филиалом или структурным подразделением другой организации, расположенной вне муниципального образования).

Практическая значимость работы заключается в применении разработанной системы поддержки принятия решения органами муниципального управления, а также научными сотрудниками в качестве методического инструментария в научно-исследовательской деятельности.

Теоретическая значимость заключается в выявлении факторов, влияющих на уровень диверсификации, а также позволяющих обосновать долю моноотрасли муниципальных образований, имеющих недостаточно данных для традиционной оценки уровня диверсификации.

1 Проблемы развития моногородов на основе диверсификации экономики

Диверсификация достаточно сложный элемент управления, особенно в моногородах. К задачам данной главы относятся изучить понятие «моногород», критерии их выделения. Дать общую характеристику понятию «диверсификация» и изучить методические основы оценки ее эффективности.

1.1 Понятие, критерии выделения и классификация моногородов

Задачами данного параграфа являются проанализировать понятие «моногород», их критерии выделения и классификации моногородов.

Понятие «моногород» не имеет единого и общего определения. Данный термин можно разбить на две составляющие основы, «монопрофильный» и «город». Приставка «моно-» в переводе с греческого означает «один». Рассмотрим понятие «город», которое является более широким по отношению к понятию «моногород». Город, как крупный населенный пункт, управляемый по особому положению и является административным, торговым, промышленным и культурным центром определяют толковые словари С. Ожегова и Н. Шведовой, Д.Н. Ушакова [39, 40]. Финансовый энциклопедический словарь отмечает наличие у города значительной численности населения, основную часть которого составляют рабочие, служащие и члены их семей, занятые вне сельскохозяйственного производства [41]. Каждый город выполняет определенную роль в деятельности региона. Все функции городов разделяют на экономическую, организационно-хозяйственную и неэкономическую.

Одними из первых исследований, посвященных моногороду «company town», были проведены Н. А. Innis [12] и Р. Н. Landis [13] в 1940-х годах. Авторы выявили в угольных моногородах США культурные изменения. А. Г. Гранберг [14] определяет понятие «моногород» как город, который

концентрирует одну отрасль хозяйства или деятельности. Словарь Мэриам-Вэбстер [36] определяет моногород как населенный пункт, зависящий от одного предприятия, которое обеспечивает все или большую услуг или функций городской жизни. Г. Грин [37] использует термин «single-company community» делает акцент на глубокую связь общества и градообразующего предприятия. Градообразующие предприятия играют важную роль в населенном пункте, так как от него зависит экономическое развитие. Такие предприятия относят к категории градообразующих, так как для большей части жителей города предприятие предоставляет рабочие места. Следовательно, материальное положение жителей города напрямую зависит от успеха работы градообразующего предприятия. Союз инвесторов России определяет моногород как город, в котором рыночные перспективы градообразующего предприятия влияют на судьбу самого города [38].

Существует проблема в формулирование критериев, с помощью которых можно отнести тот или иной город к монопрофильному. Такая необходимость возникает, когда градообразующее предприятие города находится на грани банкротства и необходимости его предоставления финансовой помощи государством. Банкротство градообразующего предприятия повлечет за собой массовые увольнения, что в свою очередь повлечет резкое ухудшение рынка труда и благополучие населения. Однако прежде, чем оказать финансовую помощь городу как городу с моноструктурой экономики, требуется на законодательном уровне закрепить критерии отнесения того или иного города к моногородам. По постановлению Правительства Российской Федерации [15] город признается моногородом, если численность населения муниципального образования более 3 тысяч человек; численность работников градообразующей организации достигала в период пяти лет, предшествующей дате утверждения перечня моногородов, 20% среднесписочной численности работников всех организаций; осуществление градообразующей

деятельности по добыче полезных ископаемых (кроме нефти и газа) и (или) производству и (или) переработке промышленной продукции.

Классификация таких городов связана с такими понятиями, как профиль, функция, специализация и основная отрасль города. П. Хаггет [28] предложил рассчитывать индекс специализации города, в результате чего можно было бы выделить промышленные, сельскохозяйственные, административные, торговые, университетские и курортные города. Аналогичным образом можно провести классификацию городов путем установления критериев определения основных направлений развития экономики города. Основную отрасль экономики было предложено выделять в [29] как отрасль промышленности, строительства, транспорта, связи, торговли или сельского хозяйства, в которой было занято не менее 15% общей численности работников последние двенадцать лет, или доля которой в объеме валового регионального продукта составляет не менее 20%.

По функциональному критерию предлагает разграничивать города Сошкин И.А. [30]: многофункциональные и монофункциональные. Ч.Д. Гаррис многофункциональным городам противопоставляет «монофункциональные» – те, которые автор характеризует определенным профилем. Таким образом, Ч.Д. Гаррис выделяет те, которые имеют моноструктуру или монопрофиль. Проблемным моментом в данном случае является пересечение терминов «функция» и «профиль» при выделении функции города с точки зрения профиля производственной деятельности. Последнее не позволит разграничить и понятия «монопрофильный» и «монофункциональный». Другим основанием классификации является профиль производственной деятельности предприятия города. Профиль города – это структура хозяйственного комплекса города, отражающая его функциональную специализацию [31]. Кузнецова Г.Ю. [42] рассматривает понятие монопрофильность как доминирование какой-либо отрасли промышленности в специализации экономической базы города.

Рассмотренное в данном разделе понятие «моногород» имеет множество определений. На основе критериев можно выделить из множество городов моногорода, наиболее оптимальный набор критериев, представлен в Постановлении Правительства РФ, а именно город признается моногородом, если численность населения муниципального образования более 3 тысяч человек; численность работников градообразующей организации достигала в период пяти лет, предшествующей дате утверждения перечня моногородов, 20% среднесписочной численности работников всех организаций; осуществление градообразующей деятельности по добыче полезных ископаемых (кроме нефти и газа) и (или) производству и (или) переработке промышленной продукции.

1.2 Методические основы оценки эффективности диверсификации

К задачам данного параграфа относятся анализ понятия «диверсификация», методов оценки ее эффективности, а именно показатели концентрации и диверсификации: доля моноотрасли, индекс на основе трех данных: о выручке, основных средствах и оплате труда предприятий моногородов.

Одной из главных проблем современной России является проблема развития и функционирование моногородов. Ведь социально-экономическое положение моногорода напрямую зависит от финансового состояния градообразующего предприятия, а такая зависимость становится угрозой существования города. Социально-экономические данные лежат в основе показателей развития моногородов, такие как уровень занятости и безработицы, доля продукции, отгруженной градообразующим предприятием, доля инвестиций в основные средства и налогов, отчисленных предприятием [1; 2].

Процесс диверсификации моногородов становится основной стратегией снижения зависимости и развития устойчивости социально-

экономического благополучия моногорода. Ансофф И. определяет, что диверсификация предполагает перераспределение ресурсов предприятия в других сферах деятельности, которые существенно отличается от действительной [32, с. 299]. Разрабатывая стратегические вопросы развития фирмы, А. Стрикленд и А. Томпсон [189] проводят систематизацию видов диверсификации, выделяя следующие стратегии диверсификации: новая отрасль; приобретение, создание новой компании и совместного предприятия. Д. Землянский и С. Лагманов рассматривают диверсификацию как основную стратегию развития моногородов [3]. Диверсификация экономики малых городов, как и в международной практики, зависит от степени вовлеченности местных и региональных властей [4, 5, 6]. Изучив 200 моногородов в Соединенных Штатах Америки Дж. Б. Аллен [16] определил жизненный цикл моногородов. Исследования последних десятилетий выявили «парадокс бедности с обилием ресурсов», характерный для данного типа населенных пунктов. [17; 18]. Ряд исследований показал особую важность диверсификации малых городов в преодолении бедности [7, 8].

На данный момент существует два подхода горизонтальный и вертикальный. Первый подход предполагает формирование инвестиционной привлекательности, расширение спектра продуктов (услуг), поддержка конкуренции на рынке, уменьшение влияния государства на экономику региона. Второй подход предполагает активное участие государства и ее политики по изменению структуры и стимуляции экономики.

Существуют разные мнения по диверсификации региональной экономики, но все авторы сходятся в одном, что диверсификация должна быть связана с ростом конкурентоспособности региона. По определению В.И. Данилова-Данильяна [9] диверсификация должна сопровождаться ростом полезного эффекта обновленной или реструктуризированной региональной хозяйственной деятельности. Для Б.Б. Дондокова [10] целью диверсификации экономики региона является обеспечение возможности реализации программ социально-экономического развития, что обеспечит

рост конкурентоспособности региона. Также Б.Б. Дондоков под диверсификацией понимает модернизацию, а не расширение видов деятельности. Т.В. Ускова и С.С. Копасова [11] диверсификация рассматривается как расширение видов деятельности с целью повышения конкурентоспособности региона и улучшения доходов населения.

Следовательно, под диверсификацией экономики региона следует понимать, как количество видов деятельности и их разнообразия в регионе, так и диверсификацию предприятий на этой территории.

Для анализа уровня и динамики диверсификации можно рассчитать следующие показатели концентрации и диверсификации на основе данных о выручке, основных средствах и оплате труда предприятий моногородов.

Доля моноотрасли (CR) рассчитывается по формуле:

$$CR_n = \sum_{i=1}^n Y_i,$$

где Y_i – доля i -го предприятия в общем объеме моногорода, n – общее число предприятий.

Индекс Херфиндаля –Хиршмана – показатель используется для оценки степени монополизации отрасли. Формула индекса Херфиндаля – Хиршмана:

$$HHI = \sum_{i=1}^n Y_i^2,$$

где Y_i – доля выручки предприятия i в общем объеме выручки предприятий города. Индекс Херфиндаля-Хиршмана принимает значения от 0 (в идеальном случае совершенной конкуренции, когда на рынке бесконечно много продавцов, каждый из которых контролирует незначительную долю рынка) до 1 (когда на рынке действует только одна фирма, производящая 100% продукции). Если считать доли рынка в процентах, индекс будет принимать значения от 0 до 10000. Чем больше значение индекса, тем выше концентрация продавцов на рынке. С 1982 года индекс Херфиндаля-Хиршмана служит основным ориентиром для реализации антимонопольной

политики США. Его основным преимуществом служит способность чутко реагировать на перераспределение долей между фирмами, действующими на рынке.

Индекс энтропии показывает среднюю долю предприятий. Формула индекса энтропии:

$$E = \sum_{i=1}^n Y_i \ln \left(\frac{1}{Y_i} \right).$$

Индекс энтропии представляет собой показатель, обратный концентрации: чем выше его значение, тем ниже концентрация предприятий на рынке. Энтропия измеряет неупорядоченность распределения долей между предприятиями рынка: чем выше показатель энтропии, тем ниже возможности продавцов влиять на рыночную цену. Для сравнения показателей энтропии на разных рынках часто используют относительный показатель энтропии.

Коэффициент Джинни:

$$G = 1 - 2 \sum_{i=1}^m x_i \text{cum} y_i + \sum_{i=1}^m x_i y_i ,$$

где x_i – доля рассматриваемого показателя моноотрасли в моногороде; y_i – доля рассматриваемого показателя не моноотрасли в моногороде; $\text{cum} y_i$ – кумулятивная доля показателя; m - число групп рассматриваемого показателя ($m=2$).

Коэффициент Розенблюта и Холла-Тайдмана:

$$R - HT = \frac{1}{2 \sum R_i q_i - 1},$$

где R_i – ранг доли рассматриваемого показателя по моноотрасли и по немоноотрасли в общей величине моногорода, q_i – доля продаж i -го предприятия на рынке.

Коэффициент вариации:

$$V = \frac{\sigma}{\alpha} \cdot 100\% ,$$

где σ - среднеквадратическое отклонение показателя; a - среднее арифметическое показателя.

Было отобрано три показателя CR, НИИ и E, они были выбраны в силу того, что сочетание показателей CR и НИИ дополняет друг друга, позволяет оценить уровень монополизации на рынке. Показатель E, так как в результате оценки происходит нормирование результатов. Мы получаем отрицательные и положительные значения, что не позволяет сделать показатель НИИ. Также совместное использование трех показателей позволяет устранить недостатки показателя выручки, такие как инфляция, «двойной счет».

Таким образом, под диверсификацией в рамках данной работы стоит понимать создание дополнительных видов экономической деятельности. Рассмотренные показатели концентрации и диверсификации планируется использовать в дальнейшем для построения математической модели.

1.3 Сравнительный анализ эффективности диверсификации моногородов РФ

В перечне монопрофильных муниципальных образований в Российской Федерации моногорода поделены на три категории.

Категория 1. Монопрофильные муниципальные образования Российской Федерации (моногорода) с наиболее сложными социально-экономическим положением (в том числе взаимосвязи с проблемами функционирования градообразующих организаций).

Категория 2. Моногорода, в которых имеются риски ухудшения социально-экономического положения.

Категория 3. Моногорода со стабильной социально-экономической ситуацией.

На рисунке 1 представлена диаграмма по количеству моногородов в каждой категории.

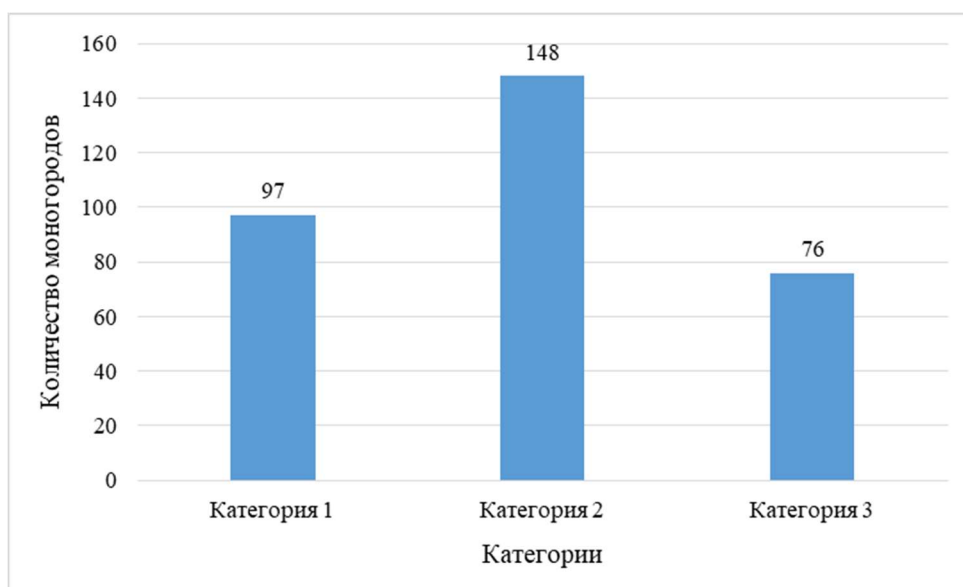


Рисунок 1 – Разделение количества моногородов по категориям

Исходя из рисунка 1, большинство моногородов принадлежат ко второй категории, то есть моногорода, в которых имеются риски ухудшения социально-экономического положения.

Сделаем разбивку моногородов по регионам и выявим регионы с наибольшей их концентрацией (рисунок 2). Из рисунка 2 можно сделать вывод, что тремя регионами с самой высокой концентрацией моногородов являются Свердловская (17 моногородов), Челябинская (16 моногородов) и Кемеровская области (24 моногорода).

Проведем оценку эффективности диверсификации в регионах с высокой концентрацией моногородов, так как высокая концентрация моногородов в пределах региона делает диверсификацию не просто стратегией развития отдельного населенного пункта, но выводит ее на региональный уровень. Для этого была сформирована выборка из 49 моногородов, расположенных в Кемеровской, Свердловской и Челябинской областях. Полная выборка моногородов трех областей составила 57 населённых пунктов.

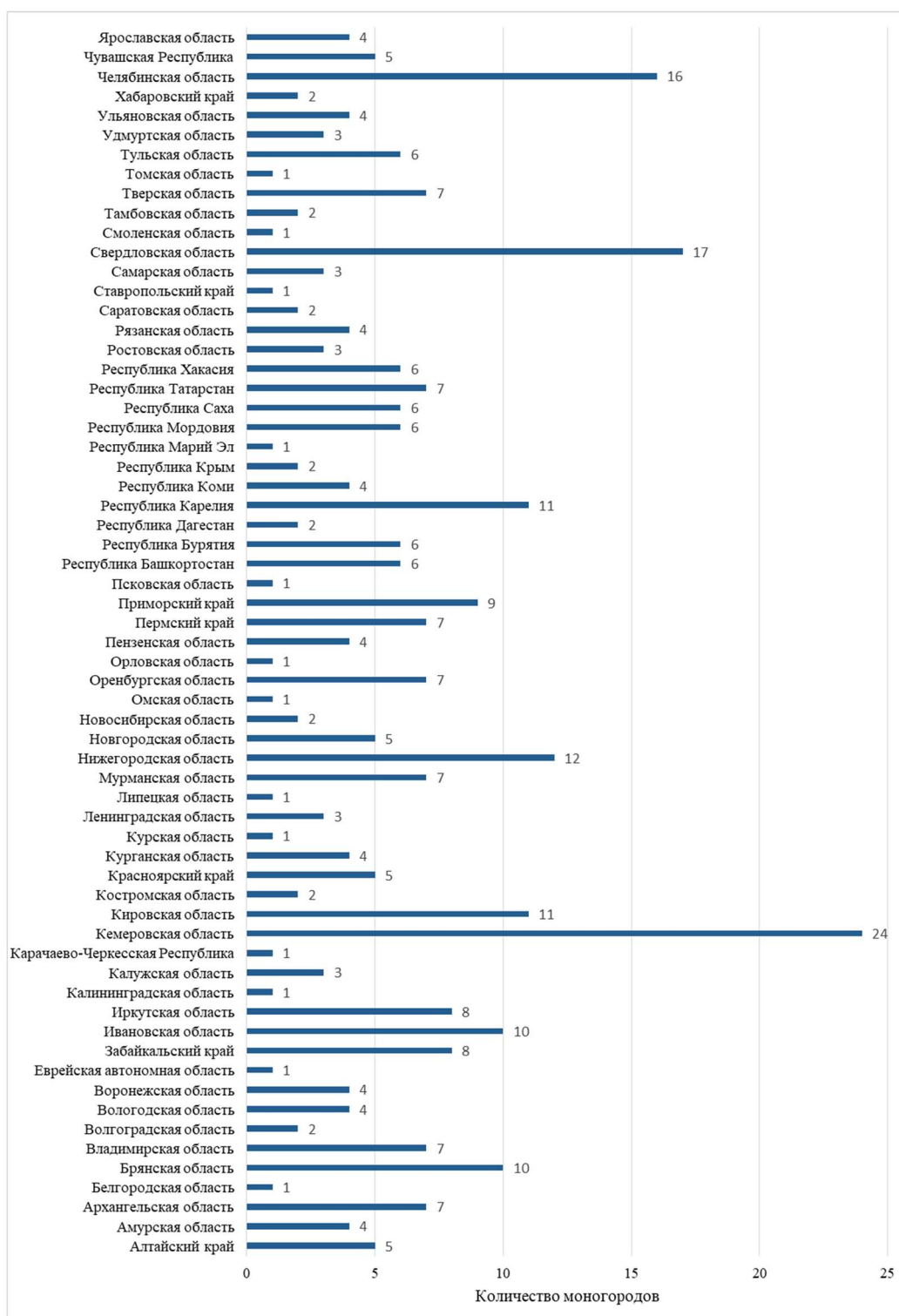


Рисунок 2 – Распределение моногородов по регионам

При этом из выборки было исключено 8 моногородов в связи с тем, что градообразующее предприятие в них представлено филиалом, что не дает возможность оценить ключевой показатель – долю монопрофильного вида деятельности в моногороде по выручке (Тайгинский, Краснобродский

городские округа, городские поселения Салаир, Белогорск, Шерегеш, Волчанск, Усть-Катавский и Нязепетровский городские округа). Итоговая выборка включила 49 населенных пунктов. Для оценки эффективности диверсификации проводилось сопоставление динамики изменения уровня диверсификации с динамикой выручки. Динамика выручки и уровня диверсификации определялась следующим образом:

$$D = \ln \frac{V_{2017}}{V_{2013}},$$

где V_{2013} и V_{2017} – исследуемый показатель (агрегированный объем выручки или значение энтропии) по отдельному моногороду за 2013 и 2017 гг. соответственно.

Диверсификация моногородов может считаться эффективной, если увеличение показателя диверсификации сопровождается ростом динамики агрегированного показателя выручки и улучшения деловой активности. В связи с чем эффективность диверсификации оценивается в двух направлениях: динамика уровня энтропии выручки с динамикой агрегированной выручки по моногороду, а также с динамикой числа вновь созданных предприятий. Первое направление отражает эффективность диверсификации во выручке, второе – по деловой активности (динамике вновь созданных предприятий).

Сопоставление динамики уровня диверсификации (энтропии) и агрегированной выручки по моногородам позволило выделить следующие моногорода, попавшие в I квадрант с положительными значениями (квадрант эффективности диверсификации):

- Мундыбаш (Кемеровская область, категория 2);
- Таштагол (Кемеровская область, категория 1);
- Верхняя Салда (Свердловская область, категория 3);
- Качканар (Свердловская область, категория 2);
- Златоуст (Челябинская область, категория 2).

Результаты представлены на рис. 3.

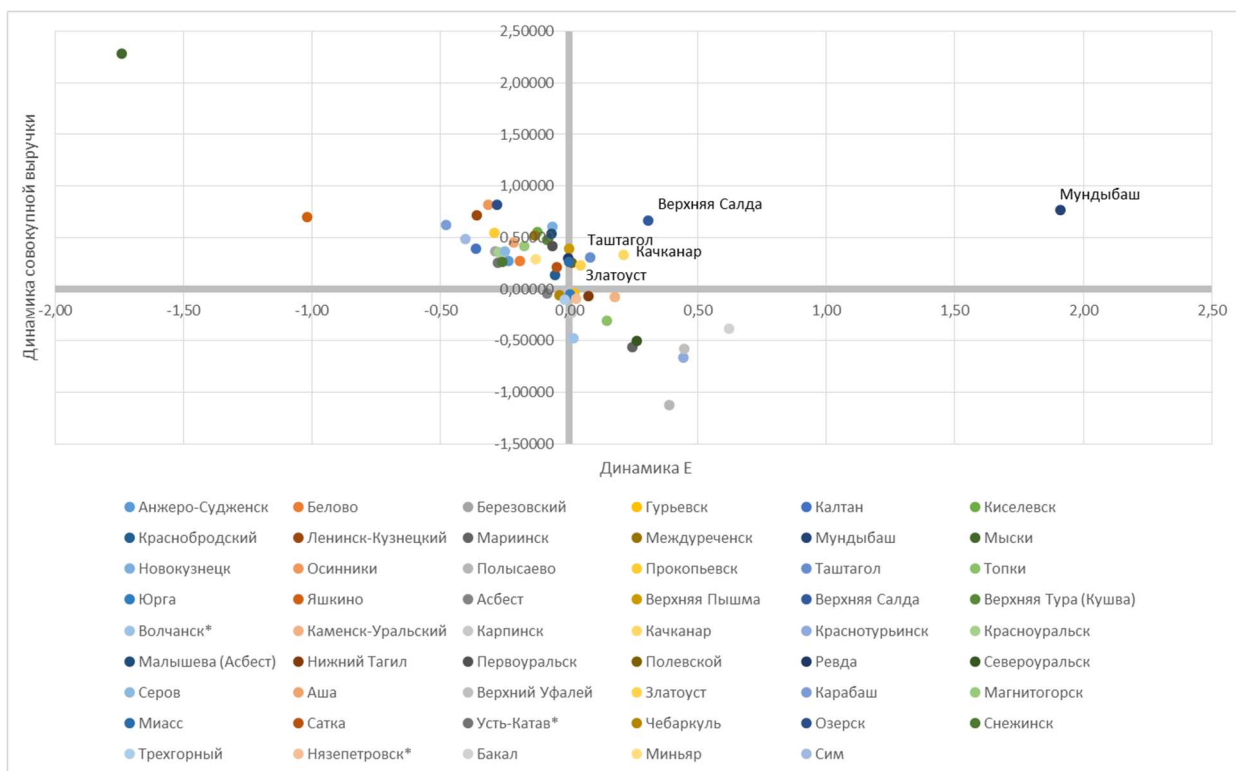


Рисунок 3- Эффективность диверсификации монопрофильных городов по энтропии и динамике совокупной выручки за 2013-2017 гг.

Принимая во внимание средний уровень инфляции 1,5 за пять лет, рассчитанный на основе статистических данных и представленный на рис. 3, ни один из моногородов не входит в квадрант эффективности, что означает отсутствие положительной динамики эффективности среди исследуемых моногородов. Единственный моногород, показавший рост агрегированной выручки выше уровня инфляции – Мыски (Кемеровская область), уровень диверсификации на максимальное среди всех моногородов значение (-1,7), что свидетельствует о том, что рост агрегированной выручки произошел за счет градообразующего предприятия, что подтверждает прирост доли выручки моноотрасли в 300 раз за пятилетний период (с 0,003 до 0,9).

Сопоставляя динамику диверсификации с динамикой деловой активности моногородов на основе роста числа вновь созданных предприятий в лидеры выходят два моногорода (рис. 4):

- Мариинск (Кемеровская область, категория 2);
- Верхний Уфалей (Челябинская область, категория 1).

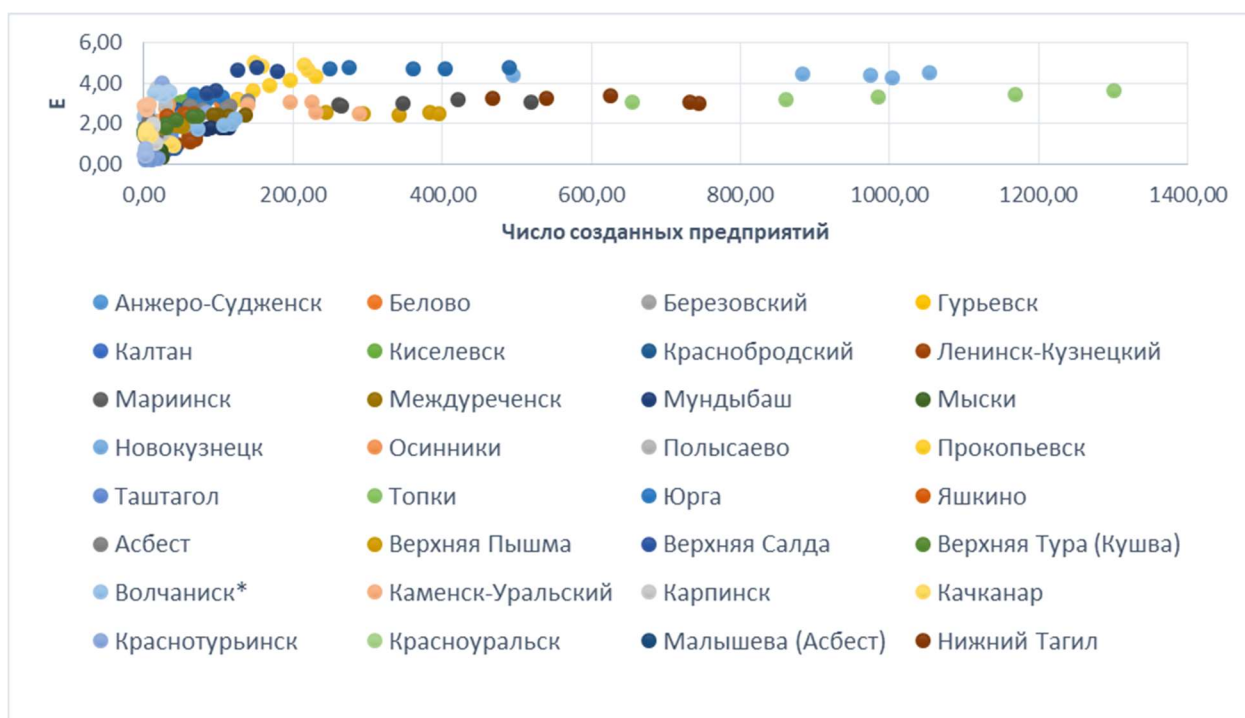


Рисунок 4 - Эффективность диверсификации монопрофильных городов по энтропии и динамике числа вновь созданных предприятий за 2013-2017 гг.

При положительно динамике деловой активности в данных моногородах заметна положительная динамика диверсификации, что свидетельствует в пользу более высокой эффективности данного процесса. Наиболее показательно то, что Верхний Уфалей был отнесен нормативными документами к первой группе риска ухудшения социально-экономического положения (с наиболее сложным положением). Так, Верхний Уфалей показывает сокращение доли выручки моноотрасли с 0,6 до 0,4 за пятилетний период, при этом растет и энтропия долей выручки предприятий в целом. Однако такое изменение связано, прежде все, с резким снижением выручки в 8 раз одного из градообразующих предприятий ОАО «Уфалей Никель», произошедшее в 2017 году (с 6,32 млрд. руб. до 0,785 млрд. руб.), что не дает основания подтвердить эффективность диверсификации экономики данного моногорода. Схожая ситуация с монопрофильными предприятиями Мариинска. АО «Мариинский ликеро-водочный завод» сократил объем производства с последующим высвобождением части персонала.

Сопоставляя полученный результат с категориями риска моногородов, можно сделать вывод о том, что среди моногородов с наилучшими показателями эффективности диверсификации моногорода, относящиеся к различным категориям риска. Это свидетельствует о том, что процесс диверсификации может протекать эффективно в моногородах с разным уровнем социально-экономического развития. Более крупные по агрегированной выручке моногорода становятся более диверсифицированными. При этом формируется один наиболее диверсифицированный центр в каждом из рассматриваемых регионов – Новокузнецк (в Кемеровской области), Магнитогорск (в Челябинской области), Нижний Тагил (в Свердловской области). Кроме того, будучи более диверсифицированными центрами, эти моногорода примерно в равной степени специализируются на тех видах деятельности, которые представлены остальными моногородами. Так, в Кемеровской области Новокузнецк специализируется, прежде всего, на металлургии, при этом добыча по объемам не уступает прочим моногородам. Остальные моногорода Кемеровской области концентрируются, в свою очередь, либо на добыче угля, либо на добыче и обработке металла, исключение составляют Яшкино (пищевая промышленность) и Тайга (транспорт). Таким образом, крупный промышленный центр выступает ядром, аккумулирующим и ретранслирующим монопрофильные виды деятельности на более мелкие моногорода, формируя кластер. Последнее свидетельствует о том, что рассматривать развитие моногородов, сконцентрированных в высокой степени в одном регионе, следует не в отдельности а в совокупности в связи с их взаимной связью и взаимной обусловленностью.

Таким образом, в разделе были рассмотрены три категории монопрофильного муниципального образования Российской Федерации. Было определено, что большинство моногородов принадлежит ко второй категории, то есть моногорода, в которых имеются риски ухудшения социально-экономического положения. Три региона с самой высокой

концентрацией моногородов являются Свердловская (17 моногородов), Челябинская (16 моногородов) и Кемеровская области (24 моногорода). Была проведена оценка эффективности диверсификации на основе роста показателя диверсификации и роста динамики агрегированного показателя выручки и роста числа вновь созданных, которая показала, что ни один из моногородов не входит в квадрант эффективности (с учетом среднего уровня инфляции 1,5), что означает отсутствие положительной динамики эффективности среди исследуемых моногородов. При положительной динамике числа вновь созданных предприятий в данных моногородах заметна положительная динамика диверсификации.

Выводы по разделу 1:

1. Рассмотрен понятийный аппарат моногорода, было определено, что наиболее оптимальный набор критериев представлен в Постановлении Правительства РФ.
2. Рассмотрены показатели концентрации и диверсификации экономики моногорода, для последующего анализа были отобраны: доля моноотрасли, индекс Херфиндаля-Хиршмана, индекс энтропии.
3. Проведена оценка эффективности уровня диверсификации экономики моногородов в регионах с высокой их концентрацией: Кемеровской, Свердловской и Челябинской областей.

2 Моделирование процесса оценки уровня диверсификации моногорода на примере Тайгинского городского округа

2.1 Общая характеристика Тайгинского городского округа

К задачам данного параграфа относятся анализ социально-экономического состояния моногорода, анализ факторов влияющих на его развитие.

Тайгинский городской округ – крупная железнодорожная станция на Транссибирской магистрали, железнодорожный узел с веткой на город Томск. Основанием для появления города Тайга являлась прокладка Великого Сибирского парового пути, первые постройки на месте города появились в 1895 году. Официально Тайга была признана населенным пунктом Томской губернии в 1896 году [34]. В 1911 году Тайга из поселка была обращена в городское поселение, численность населения которого составляла 10100 человек. Тайгинский городской округ образован 17 декабря 2004 г.

Перепись населения 1920 года дала более четкое представление о жителях Тайги. Численность населения составила 12035 человек, железнодорожники с семьями составляли 73% от общего населения. Также работали помимо железнодорожных предприятий 4 кузницы, 2 сапожные, 2 столярные и жестяные мастерские, кирпичный и пивоваренный заводы, часовая мастерская – всего 21 предприятие, на которых были заняты 127 человек. [34]. Таким образом, можно сделать вывод, что город существовал для обслуживания и функционирования железнодорожной станции. По переписи населения 1937 года численность населения составляла уже 31,6 тысячи человек. На рисунке 5 показано изменение численности населения с 1920 по 2019 год.

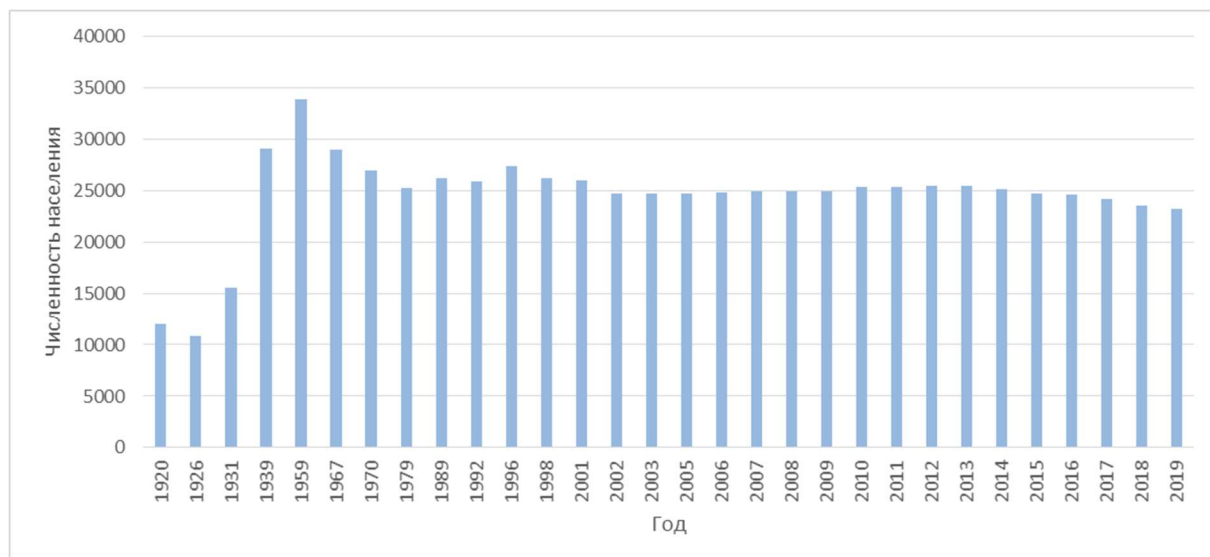


Рисунок 5 – Динамика численности населения

До 1959 года население заметно растет. Послевоенные пятилетки примечательны для города его бурным развитием и трудом железнодорожников. Появились новые жилые кварталы, промышленные предприятия.

Для рассмотрения разброса предприятий Тайгинского городского округа по выручке, основным средствам и оплате труда по годам был проведен дисперсионный анализ. На рисунке 6 представлена диаграмма размаха выручки предприятий Тайгинского городского округа. На рисунке 8 – диаграмма размаха предприятий по основным средствам. На рисунке 10 – диаграмма размаха предприятий по оплате труда.

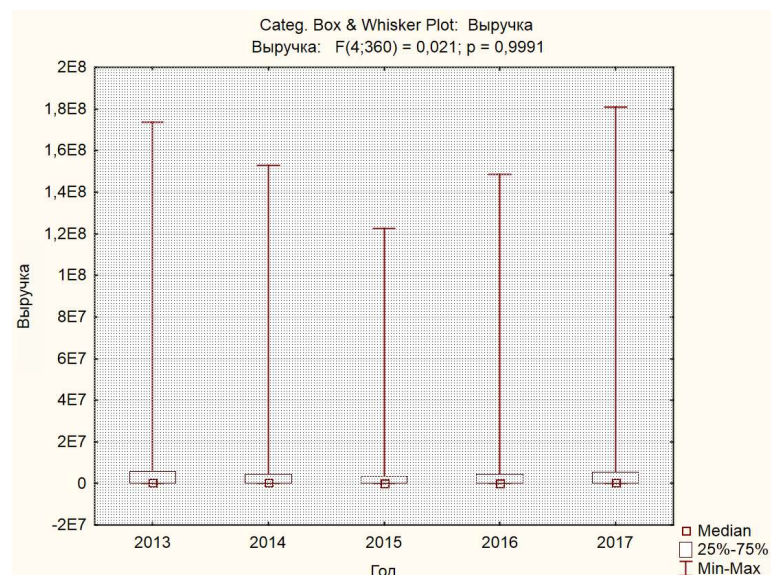


Рисунок 6 – Диаграмма размаха выручки предприятий Тайгинского городского округа

Описательная статистика о разбросе выручки представлена на рисунке 7.

Год	Выручка Means	Выручка Std.Dev.	Выручка Variance	Выручка Minimum	Выручка Maximum
2013	8830233	24148914	5,831700E+14	0,00	173619000
2014	9417014	25689437	6,599472E+14	0,00	152850000
2015	8735466	22638931	5,125212E+14	0,00	122342000
2016	9459507	25353195	6,427845E+14	0,00	148441000
2017	8549945	24213217	5,862799E+14	0,00	180926000

Рисунок 7 – Описательная статистика разброса выручки предприятий Тайгинского городского округа

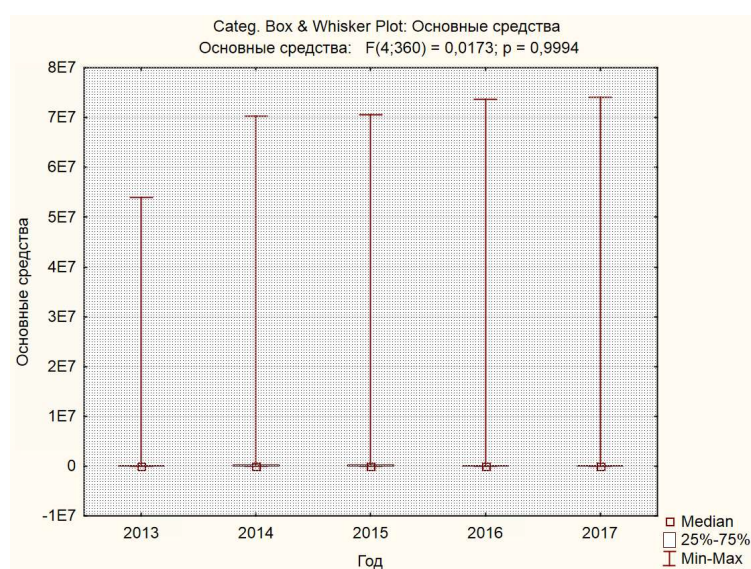


Рисунок 8 – Диаграмма размаха основных средств предприятий Тайгинского городского округа

Описательная статистика о разбросе основных средств представлена на рисунке 9.

Год	Основные средства Means	Основные средства Std.Dev.	Основные средства Variance	Основные средства Minimum	Основные средства Maximum
2013	2665890	10023257	1,004657E+14	0,00	53942000
2014	3124411	11530956	1,329629E+14	0,00	70273000
2015	2750644	10994925	1,208884E+14	0,00	70516000
2016	2916918	12283237	1,508779E+14	0,00	73622000
2017	2832562	12141675	1,474203E+14	0,00	74128000

Рисунок 9 - Описательная статистика разброса основных средств предприятий Тайгинского городского округа

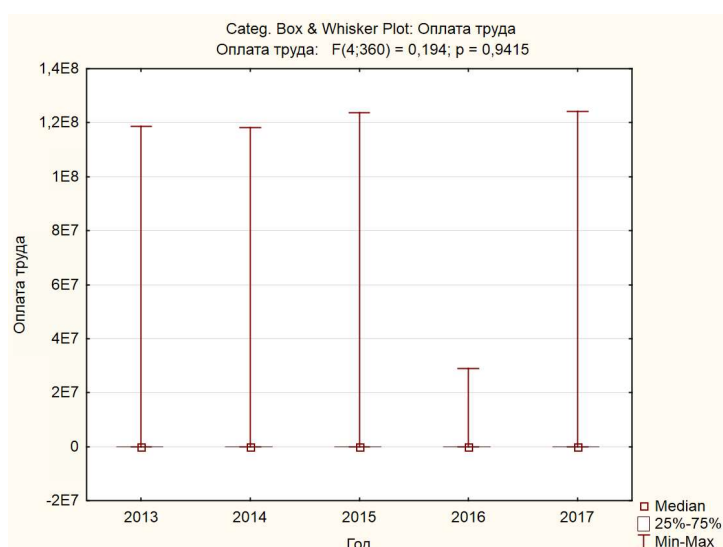


Рисунок 10 – Диаграмма размаха оплаты труда предприятий Тайгинского городского округа

Описательная статистика представлена на рисунке 11.

Год	Оплата труда Means	Оплата труда Std.Dev.	Оплата труда Variance	Оплата труда Minimum	Оплата труда Maximum
2013	2415808	14376390	2,066806E+14	0,00	118549000
2014	2387699	14230818	2,025162E+14	0,00	118205000
2015	2005849	14690625	2,158145E+14	0,00	123770000
2016	763384	4583243	2,100612E+13	0,00	28920000
2017	2099781	14829494	2,199139E+14	0,00	124130000

Рисунок 11 - Описательная статистика разброса оплаты труда предприятий Тайгинского городского округа

Проверка гипотезы о равенстве средних принимается, так как $p > 0,05$, что говорит о том, что год на изменение выручки, основных средств и оплаты

труда не влияет. Наблюдается достаточно большой разброс данных по всем трем показателям

Реструктуризация железнодорожного транспорта, начавшаяся с 1997 года, негативно отразилась на экономике города. Все предприятия были лишены статуса юридического лица, за период с 1998-2015 годы численность работающих на градообразующем предприятии сократилась почти на 1800 человек (31,8% от максимального уровня 2006 года). На данный момент 51% населения занято на предприятиях ОАО «РЖД», квалификационный состав населения ориентирован на потребностях данных предприятий. Экономика Тайгинского городского округа является слабо диверсифицированной, доход бюджета сильно зависит от деятельности предприятий железнодорожного транспорта.

Распоряжением Правительства РФ от 29 июля 2014 года № 1398-р «Об утверждении перечня моногородов» городской округ включён в категорию «Монопрофильные муниципальные образования Российской Федерации (моногорода), в которых имеются риски ухудшения социально-экономического положения». Население на 1 января 2019 года составляет 24948 человек. Городской округ и город областного подчинения включают населённые пункты: Тайга, Таежный, Сураново, Пихтач, Кузель, Кедровый. Градообразующим предприятием является ОАО «РЖД», среднесписочная численность работников которой составляет 3291 человек.

Исторически Тайгинский городской округ связан с железнодорожной деятельностью, сейчас его градообразующим предприятием является ОАО «Российские железные дороги» и включен в перечень монопрофильных муниципальных образований РФ. Был проведен дисперсионный анализ по трем показателям предприятий по выручке, основным средствам и оплате труда, наблюдался достаточно высокий размах данных.

2.2 Систематизация методов машинного обучения для поддержки принятия решения об уровне и направлениях диверсификации моногорода

К задачам текущего раздела относится описание алгоритмов методов машинного обучения, которые применялись для моделирования доли моноотрасли Тайгинского городского округа.

Регрессия

Регрессия является одним из наиболее известных и понятных алгоритмов в статистике и машинном обучении.

Допустим, мы хотим иметь систему, которая может предсказать цену подержанного автомобиля. Входные данные - это атрибуты автомобиля: марка, год выпуска, мощность двигателя, пробег и другая информация, которая, по нашему мнению, влияет на стоимость автомобиля. Выход - это цена автомобиля. Такие задачи, где выходное число-это задачи регрессии. Пусть X обозначает атрибуты автомобиля, а Y -цену автомобиля. Опять же, исследуя прошлые транзакции, мы можем собрать обучающие данные, и программа машинного обучения приспособливает функцию к этим данным, чтобы узнать Y как функцию X [19]. Например, регрессионная модель выручки телекоммуникаций от числа зарегистрированных предприятий:

$$Y = 0,747 \cdot X_1,$$

где Y – выручка телекоммуникаций, X_1 – число зарегистрированных предприятий.

Как регрессия, так и классификация являются контролируемыми задачами обучения, где вход - X , выход - Y , и задача состоит в том, чтобы изучить отображение от входа к выходу. Подход в машинном обучении заключается в том, что мы предполагаем модель, определенную с точностью до набора параметров:

$$y = g(x | \theta),$$

где $g(\cdot)$ - модель, а θ -ее параметры. Y - это число в регрессии и код класса (например, 0/1) в случае классификации. $g(\cdot)$ - регрессионная функция или в классификации-дискриминантная функция, разделяющая экземпляры различных классов. На рис. 12 представлены виды регрессии линейная и нелинейная.

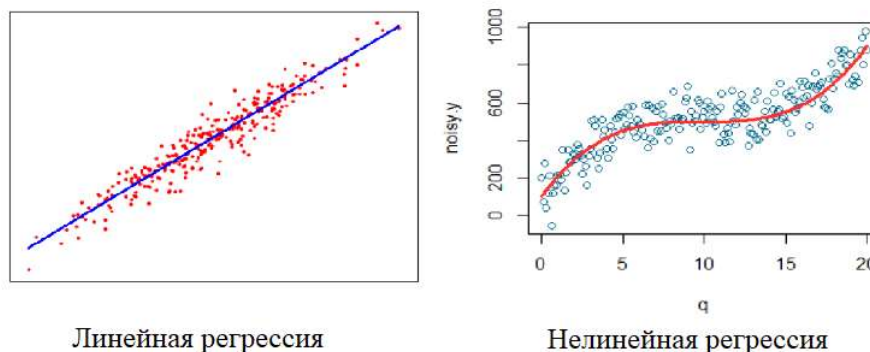


Рисунок 12 – Виды регрессии

Программа машинного обучения оптимизирует параметры θ таким образом, чтобы ошибка аппроксимации была минимизирована, то есть наши оценки максимально приближены к правильным значениям, заданным в обучающем наборе.

В тех случаях, когда линейная модель слишком ограничительна, можно использовать, например, квадратичную функцию:

$$y = w_0 + w_1x + w_2x^2$$

или полином более высокого порядка, или любая другая нелинейная функция входного сигнала, на этот раз оптимизирующая его параметры для наилучшего соответствия. Простейшая и наиболее известная экономико-математическая модель теории производства в постановочном плане и с точки зрения применяемых математических методов является производственная функция Кобба-Дугласа:

$$y = f(x_1, x_2) = A \cdot L_1^\alpha \cdot K_2^\beta,$$

где L – затраты труда, K – затраты капитала.

Алгоритм случайный лес

Данный алгоритм машинного обучения, который был предложен Лео Брейманом и Адель Катлер заключается в использовании ансамбля решающих деревьев [20]. Алгоритм включает в себе две идеи: метод бэггинга Бреймана и метод случайных подпространств Tin Kam Ho. Данный алгоритм применяется для задач классификации, регрессии и кластеризации.

Решающие деревья – хорошее семейство базовых классификаторов для бэггинга. Метод случайных подпространств снижает коррелированность между деревьями и позволяет избегать переобучения. Алгоритм построения случайного леса для N деревьев выглядит следующим образом [21]:

- 1) Для каждого $n = 1, \dots, N$: генерировать выборку X_n с помощью бутстрэпа;
- 2) Построить решающее дерево b_n из выборке X_n :
 - по заданному критерию выбирается лучший признак и по нему делаем разбиение в дереве и так до исчерпания выборки;
 - дерево строится, пока в каждом листе не более n_{\min} объектов или пока дерево не достигнет определенной высоты;
 - при каждом разбиении выбирается m случайных признаков из p исходных, среди которых ищется оптимальное разделение.

Рекомендуется для задач регрессии брать $m = \frac{p}{3}$.

- 3) Итоговый классификатор $a(x) = \frac{1}{N} \sum_{i=1}^N b_i(x)$.

Метод ближайших соседей

Алгоритм k -ближайших соседей (k -NN) является непараметрическим методом, используемым для классификации и регрессии [22] в обоих случаях входные данные состоят из k ближайших обучающих примеров в пространстве признаков. Результат зависит от того, используется ли k -NN для классификации или регрессии.

В классификации k-NN выводом является принадлежность к классу. Объект классифицируется по множеству голосов его соседей, причем объект присваивается классу, наиболее распространенному среди его k ближайших соседей (k-положительное целое число, обычно маленькое). Если $k = 1$, то объект просто присваивается классу этого единственного ближайшего соседа.

В регрессии k-NN выходным значением является значение свойства для объекта. Это значение является средним значением значений k ближайших соседей.

Метод ближайших соседей - это тип обучения на основе экземпляра, или ленивого обучения, где функция только аппроксимируется локально, и все вычисления откладываются до оценки функции.

Как для классификации, так и для регрессии полезным методом может быть присвоение весов вкладам соседей, так что более близкие соседи вносят больший вклад в среднее значение, чем более отдаленные. Например, общая схема взвешивания заключается в придании каждому соседу веса $1/d$, где d-расстояние до соседа.

Соседи берутся из набора объектов, для которых известен класс (для классификации k-NN) или значение свойства объекта (для регрессии k-NN). Это можно рассматривать как обучающий набор для алгоритма, хотя явный этап обучения не требуется.

Особенностью алгоритма k-NN является то, что он чувствителен к локальной структуре данных.

Алгоритм метода выглядит следующим образом:

- Примеры обучения - это векторы в многомерном объектном пространстве, каждый из которых имеет метку класса. Этап обучения алгоритма состоит только из хранения векторов признаков и меток классов обучающих выборок.

- На этапе классификации k является определяемой пользователем константой, а немаркированный вектор (запрос или тестовая точка) классифицируется путем назначения метки, которая является наиболее распространенной среди K обучающих выборок, ближайших к этой точке запроса. Обычно используемая метрика расстояния для непрерывных переменных - это евклидово расстояние. Для дискретных переменных, например для классификации текста, можно использовать другую метрику, такую как метрику перекрытия (или расстояние Хэмминга). Например, в контексте данных микрочипов экспрессии генов в качестве метрики использовался k -NN с коэффициентами корреляции, такими как Пирсон и Спирмен [23]. Часто точность классификации k -NN может быть значительно улучшена, если метрика расстояния изучается с помощью специализированных алгоритмов, таких как анализ ближайших соседей с большим запасом или анализ компонентов окрестностей.

Недостаток базовой классификации "голосования большинством голосов" возникает тогда, когда распределение классов искажено. То есть примеры более частого класса, как правило, доминируют в предсказании нового примера, потому что они, как правило, распространены среди k ближайших соседей из-за их большого числа [24]. Одним из способов преодоления этой проблемы является взвешивание классификации с учетом расстояния от тестовой точки до каждого из ее k ближайших соседей. Класс (или значение, в задачах регрессии) каждой из k ближайших точек умножается на вес, пропорциональный обратному расстоянию от этой точки до тестовой точки. Другой способ преодоления перекоса-это абстракция в представлении данных. Например, в самоорганизующейся карте каждый узел является представителем (центром) кластера сходных точек, независимо от их плотности в исходных данных обучения k -NN будут применены к самоорганизующейся карте.

Нейронная сеть прямого распространения

Нейронная сеть – это математическая модель, ее программное или аппаратное воплощение, построенная по принципу организации и функционирования биологических нейронных сетей [25]. Первой попыткой смоделировать процессы протекающие в мозге были нейронные сети У. Маккалова и У. Питтса [26]. Разработав алгоритмы обучения, модели стали использовать в задачах прогнозирования, для распознавания образов, в задачах управления и др.

В нейронной сети прямого распространения все связи направлены строго от входных нейронов к выходным (рис. 13). Нейронные сети часто описываются в виде слоенного торта, где каждый слой состоит из входных, скрытых или выходных клеток. Клетки одного слоя не связаны между собой, а соседние слои обычно полностью связаны.

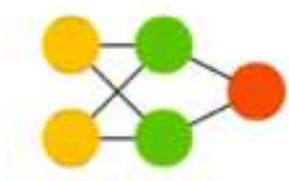


Рисунок 13 – Схема нейронной сети прямого распространения. (Желтым цветом обозначены входные нейроны, зеленым скрытые нейроны, красным – выходной нейрон)

Простейшая нейронная сеть имеет две входных клетки и одну выходную и может использоваться в качестве модели логических вентилей. Нейронная сеть прямого распространения обычно обучается с использованием метода обратного распространения ошибки, при котором сеть получает множества входных и выходных данных. Этот процесс называется обучением с учителем, и он отличается от обучения без учителя тем, что во втором случае сеть независимо формирует набор выходных данных. Вышеуказанная ошибка является разницей между входом и выходом. Если в сети имеется достаточное количество скрытых нейронов,

она теоретически способна смоделировать взаимодействие между входным и выходными данными.

Таким образом, были рассмотрены следующие методы машинного обучения: регрессия, случайный лес, метод ближайших соседей и нейронные сети прямого распространения.

2.3 Обоснование языка программирования для разработки модели машинного обучения

В качестве языка программирования был выбран язык Python. Python является самым популярным высокоуровневым языком программирования с динамической семантикой. Язык Python часто используется в машинном обучении. Так как для таких задач как собирать, систематизировать и анализировать данные, а затем на основе полученной информации создавать алгоритмы для искусственного интеллекта данный язык подходит лучше всего. К тому же у него отличная производительность при обработке данных.

Одна из основных причин, почему Python используется для машинного обучения состоит в том, что у него есть множество фреймворков, которые упрощают процесс написания кода и сокращают время на разработку. В научных расчетах используется Numpy, в продвинутых вычислениях — SciPy, в извлечении и анализе данных — SciKit-Learn. Эти библиотеки работают в таких фреймворках, как TensorFlow, CNTK и Apache Spark.

Можно выделить следующие преимущества языка программирования Python:

- Понятность. Язык программирования Python является самым высокоуровневым языком. Простой синтаксис позволяет тестировать сложные алгоритмы с минимальной затратой временных ресурсов на их реализацию.

- Обширная поддержка. Множество полезных ресурсов и качественная документация для программистов.

Таким образом, перечисленные факторы показывают, почему язык программирования Python активно используется для машинного обучения. Перечисленные выше преимущества стали основополагающими для выбора именно этого языка программирования, что позволило быстро и качественно осуществить поставленные задачи.

Выводы по разделу 2:

1. Дана общая характеристика объекта исследования. На основе дисперсионного анализа финансовых показателей предприятий Тайгинского городского округа сделан вывод о том, что год на изменение выручки, основных средств и оплаты труда не влияет. Наблюдается достаточно большой разброс данных по всем трем показателям.
2. Описаны алгоритмы метода машинного обучения и обосновано их применение для построения модели предсказания доли моноотрасли.
3. Для осуществления выбранных алгоритмов и работы с полученной базой данных был выбран язык программирования Python.

3 Разработка модели машинного обучения для предсказания значения доли моноотрасли на примере Тайгинского городского округа

В данной главе рассматривается информационная система и модель машинного обучения как элемент поддержки принятия управленческого решения. Рассматриваются выгруженные данные и их подготовка для их последующего анализа. Описывается получение моделей машинного обучения для предсказания доли моноотрасли Тайгинского городского округа на языке программирования Python.

3.1 Экономические информационные системы как элемент поддержки принятия управленческого решения

Задачей данного раздела является рассмотреть экономическую информационную систему и ее роль в принятии управленческого решения.

Экономическая информационная система – это система, предназначенная для сбора, хранения, поиска, обновления, обработки и выдачи информации во времени о деятельности экономического объекта по запросам пользователей.

Экономическая информационная система представляет - это совокупность внутренних и внешних потоков прямой и обратной информации об экономическом объекте, методах, средствах, специалистах, вовлеченных в процессе обработки информации и принятии управленческих решений. Важнейшими функциями экономической информационной системы являются: учет, анализ, прогнозирование и планирование экономических процессов.

В экономике с учетом сферы применения выделяются следующие информационные системы:

- банковские информационные системы;
- страховые информационные системы;

- статистические информационные системы;
- информационные системы фондового рынка;
- налоговые информационные системы;
- бухгалтерские информационные системы и т.д.

Конечной целью функционирования экономической информационной системы является эффективное управление экономической системой (экономическим объектом): промышленным предприятием, торговой организацией, коммерческим банком, государственным учреждением и т.д.

Рассмотрим место экономической информационной системы в управлении экономическими объектами, схема представлена на рисунке 14.

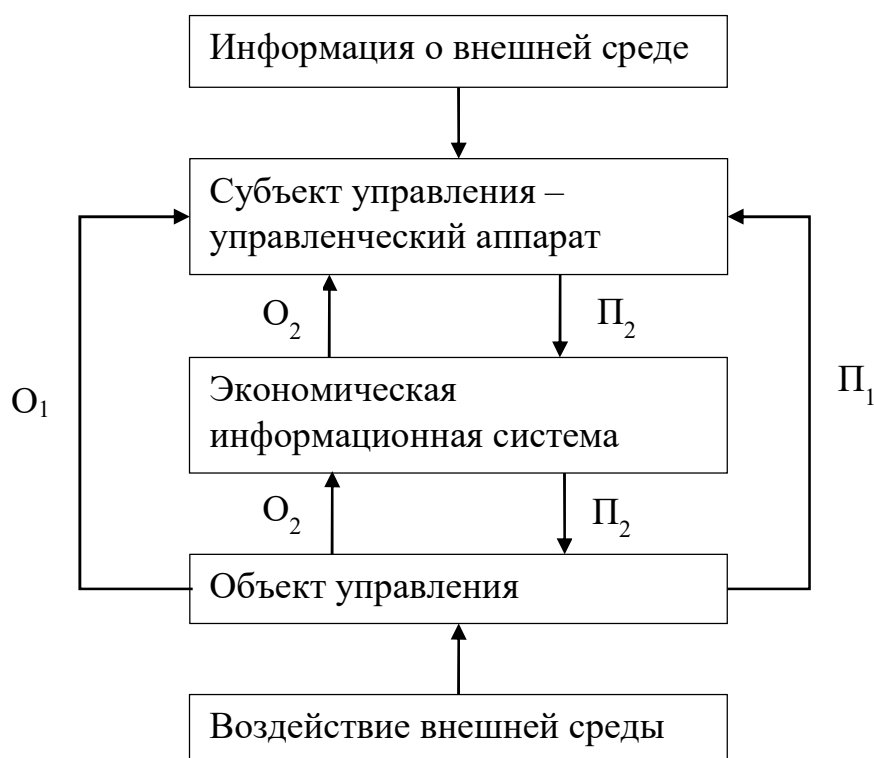


Рисунок 14 – Место экономических информационных систем в контуре системы управления

На рисунке 14 Π_1 – не формализуемая часть директивной информации, направляемой от управленческого аппарата к объекту управления (прямая связь), O_1 – не формализуемая часть отчётной информации о выполнении принятых решений, направляемая в обратном

направлении, то есть, от объекта управления к субъекту управления (обратная связь); O_2 – формализованная информация, которую можно систематизировать и обрабатывать с помощью компьютера; Π_2 - директивная информация, которая может быть соответствующим образом переработана и передана объекту управления.

Современное программное обеспечение корпораций, фирм формата информационных систем учёта и управления всё чаще ассоциируют с понятием «экономическая модель». Экономическое моделирование предполагает использование информационных технологий автоматизации учёта и поддержки принятия управленческих решений. Понятие «модель» является важнейшей категорией теории познания. На идее моделирования основаны как теоретические методы, связанные с использованием абстрактных моделей, так и экспериментальные, предусматривающие практические приложения. Термин «моделирование» используют в экономике по-разному. Так, при настройке информационной системы учёта и управления оперируют понятием, информационная модель, определяющим, прежде всего, форматы хранения данных. Информационное моделирование, как правило, дополняется характеризуемым сложными расчётами математическим моделированием. «Экономическая модель», с одной стороны, объединяет указанные два понятия, с другой – отражает и предметную сущность решаемых управленческих проблем.

Из всего разнообразия видов информационных систем можно выделить систему поддержки принятия решений. Данная система позволяет обслуживать частично структурированные задачи, результаты которых сложно спрогнозировать заранее. Они имеют мощный аналитический аппарат с несколькими математическими моделями. Система поддержки принятия решений (СППР) (англ. Decision Support System, DSS) — это компьютерная автоматизированная система, целью которой является помощь людям, принимающим решение в сложных условиях для полного и объективного анализа предметной деятельности. Данные системы возникли в

результате слияния управленческих информационных систем и систем управления базами данных.

Система поддержки принятия решений использует различные методы для анализа и выработки предложений. Это могут быть: поиск информации, интеллектуальный анализ данных, поиск знаний в базах данных, логическое обоснование имитационное моделирование, генетические алгоритмы, нейронные сети и др. Некоторые из этих методов были разработаны в рамках искусственного интеллекта. Если система поддержки принятия решений основана на методах искусственного интеллекта, то речь идет об интеллектуальной СППР, или ИСППР. Классы систем близкие к СППР — это экспертные системы и автоматизированные системы управления.

Рассмотрев экономические информационные системы и их место в принятии управленческих решений, можно сказать о том, что это актуальная сфера информационных приложений. Применение данных систем обеспечивает повышение эффективности работы управляющих.

3.2 Модель машинного обучения как элемент системы поддержки принятия управленческого решения

Чтобы решить задачу на компьютере, нам нужен алгоритм. Алгоритм - это последовательность действий, которые должны быть выполнены для преобразования входных данных в выход. Входные данные — это набор чисел, а выход - их упорядоченный список. Для одной и той же задачи могут существовать различные алгоритмы, и мы можем быть заинтересованы выходные. Например, можно разработать алгоритм сортировки. Вход - это набор в поиске наиболее эффективного, требующего наименьшего количества действий или памяти, или того и другого.

Применение методов машинного обучения к большим базам данных называется data mining. Можно провести аналогию с шахтой, из шахты извлекается большой объем земли и сырья, что при обработке приводит к

небольшому количеству очень ценного материала; аналогично, в интеллектуальном анализе данных большой объем данных обрабатывается для построения простой модели с ценным использованием, например, обладая высокой прогностической точностью. Области его применения весьма обширны, например, розничная торговля, в сфере финансов банки анализируют свои прошлые данные для построения моделей, используемых в кредитных заявках, выявлении мошенничества и на фондовом рынке. В производстве обучающие модели используются для оптимизации, контроля и устранения неполадок. В медицине обучающие модели используются для медицинской диагностики. В телекоммуникациях модели анализируют оптимизацию сети и максимизацию качества обслуживания. В таких науках как физика, астрономия и биология большие объемы данных могут быть достаточно быстро проанализированы только компьютерами. Всемирная паутина огромна, она постоянно растет, и поиск нужной информации не может осуществляться вручную.

Но машинное обучение - это не только проблема базы данных; это также часть искусственного интеллекта. Чтобы быть разумной, система, находящаяся в изменяющейся среде, должна обладать способностью к обучению. Если система может учиться и адаптироваться к таким изменениям, то проектировщику системы не нужно предвидеть и предлагать решения для всех возможных ситуаций. Машинное обучение также помогает нам находить решения многих проблем в области зрения, распознавания речи и робототехники.

Машинное обучение - это программирование компьютеров для оптимизации критерия производительности с использованием примеров данных или прошлого опыта. У нас есть модель, определенная для некоторых параметров, и обучение - это выполнение компьютерной программы для оптимизации параметров модели с использованием обучающих данных или прошлого опыта. Модель может быть предсказательной, чтобы делать

прогнозы в будущем, или описательной, чтобы получать знания из данных, или и то, и другое [19].

Машинное обучение использует теорию статистики при построении математических моделей, поскольку основная задача состоит в том, чтобы сделать вывод из выборки. Роль информатики двояка: во-первых, в обучении нам нужны эффективные алгоритмы для решения задачи оптимизации, а также для хранения и обработки огромного объема данных, которыми мы обычно располагаем. Во-вторых, как только модель изучена, ее представление и алгоритмическое решение для вывода также должны быть эффективными. В некоторых приложениях эффективность алгоритма обучения или вывода, а именно его пространственно-временная сложность, может быть столь же важна, как и его предсказательная точность.

3.3 Система поддержки оценки и принятия решения о диверсификации моногорода

Задачами данного раздела заключаются в подготовке базы данных по трем регионам с высокой концентрацией моногородов, а именно Кемеровской, Челябинской и Свердловской областей, с расчетом коэффициентов концентрации и диверсификации моногородов, получении модели машинного обучения и проверки ее качества для возможности предсказания доли «скрытой» моноотрасли Тайгинского городского округа с помощью языка программирования Python.

Для расчета показателей моногородов использовались данные финансовой отчетности предприятий, представленные в аналитической системе «СПАРК-Интерфакс». Была сформирована база данных, состоящая из панельных данных за период 2013-2017 гг. трех регионах Кемеровской, Свердловской и Челябинской областях.

Сбор данных о моногородах ограничен временными рамками, поскольку официальные данные статистики предоставляются в основном с

2006 по 2009 гг. Изучение концентрации видов экономической деятельности в течении нескольких лет мешает смена классификатора ОКВЭД, имевшее место быть в 2016 г., а также неравномерное число предприятий, представлявших отчётность в органы статистики. Так, если в период 1999-2005 гг. в информационной базе СПАРК по Анжеро-Судженску представлено лишь 22 предприятия, то в период 2006-2010 гг. – 39, 2011-2015 гг. – 152 предприятия. Снижение данных ограничений может быть связано как с выделением базового числа предприятий, так и вновь появляющихся, что открывает новые направления исследования в области динамики развития моногородов.

В первоначальной базе по Кемеровской области было 6785 предприятий, которые находятся и осуществляют свою деятельность в 24 моногородах Кемеровской области. Далее предприятия, которые не имели выручки в период 2013-2017 гг. были исключены из выборки, а также вертикально-интегрированные компании, осуществляющие деятельность в нескольких моногородах (ОАО «УК Кузбассразрезуголь» - Кемерово, Калтан, Краснобродский и др.). Также были исключены Белогорск, Салаир, Тайга, Шерегеш. Финальная выборка содержала 2557 компаний, включая 30 монопрофильных предприятий (МПП), Общая характеристика выборки на примере 2017 года представлена в таблице 1.

Таблица 1 - Общая характеристика выборки предприятий по моногородам Кемеровской области в 2017 г.

№	Моногород	Число пр.	Число МПП	CR	ННІ	Е	Градообразующее предприятие
1	Анжеро-Судженск	69	2	0,00015	0,489	1,669	ОАО «Шахтоуправление Анжерское», ОАО «Анжерский машиностроительный завод»
2	Белово	166	4	0,613	0,195	2,636	ООО «Шахта Грамотеинская», ООО «Шахта Чертинская-Коксовая», ООО «Шахта Листвяжная», «Бачатский угольный разрез»
3	Березовский	49	1	0,865	0,381	1,571	ОАО «Березовский сыродельный комбинат»

Продолжение таблицы 1

4	Гурьевск	35	1	0,296	0,255	1,886	ОАО «Гурьевский металлургический завод»
5	Калтан	38	2	0,642	0,477	1,192	Шахта «Алардинская», Южно-Кузбасская ГРЭС
6	Киселевск	127	6	0,773	0,124	2,823	ООО «Шахта № 12», ООО «Шахта Киселевская», ООО «Разрез Киселевский», ООО «Участок Коксовый», ОАО «Разрез Октябрьский», ОАО «Поляны»
7	Краснобродский	19	1	0,749	0,094	2,671	ООО «Краснобродский угольный разрез»
8	Ленинск-Кузнецкий	145	1	0,866	0,722	0,971	ОАО «СУЭК-Кузбасс»
9	Мариинск	35	1	0,870	0,304	1,972	АО «Мариинский ликёро-водочный завод»
10	Междуреченск	162	1	0,927	0,157	0,250	ЗАО «Распадская угольная компания»
11	Мундыбаш	9	1	0,006	0,258	1,616	Мундыбашская обогатительная фабрика
12	Мыски	49	1	0,877	0,771	0,746	ОАО «Южный Кузбасс»
13	Новокузнецк	1149	1	0,220	0,058	4,262	ОАО «ЗСМК», включая ОАО «Новокузнецкий металлургический комбинат»
14	Осинники	76	1	0,485	0,270	2,131	Шахта «Заречная»
15	Полысаево	33	1	0,698	0,494	1,565	Шахта «Заречная»
16	Прокопьевск	230	1	0,636	0,149	3,215	ООО «Шахта имени Дзержинского»
17	Таштагол	39	1	0,803	0,532	1,357	Таштагольский филиал ОАО «Евразруда»
18	Топки	29	1	0,8	0,642	1,119	ООО «Топкинский цемент»
19	Юрга	85	1	0,133	0,134	3,424	ООО «Юргинский машиностроительный завод»
20	Яшкино	13	1	0,961	0,917	0,264	ООО «КДВ Яшкино»
	ИТОГО	2557	30				

В первоначальной базе Челябинской области количество предприятий составляло 49438, которые находятся и осуществляют свою деятельность в 16 моногородах. Предприятия, которые не имели выручки в период 2013-2017 гг были исключены, также был исключен Усть-Катав и Нязепетровск. Общая характеристика выборки представлена в таблице 2.

Таблица 2 - Общая характеристика выборки предприятий по моногородам Челябинской области в 2017 г.

№	Моногород	Число МПП	CR	HHI	E	Градообразующее предприятие
1	Аша	1	0,672	0,262	1,614	ОАО «Ашинский металлургический завод»
2	Верхний Уфалей	1	0,294	0,130	2,926	ОАО «Уфалейникель»
3	Златоуст	1	0,08	0,03	4,851	ОАО «Златоустовский машиностроительный завод»
4	Карабаш	1	0,972	0,946	0,196	ЗАО «Карабашмедь»
5	Магнитогорск	1	0,561	0,290	3,081	ОАО «Магнитогорский металлургический комбинат»
6	Миасс	1	0,324	0,052	4,708	Автомобильный завод «УРАЛ»
7	Сатка	1	0,650	0,243	2,359	Комбинат «Магнезит»
8	Чебаркуль	1	0,622	0,409	1,793	ОАО «Уральская кузница»
9	Озерск	1	0,37	0,147	3,484	Производственное объединение «Маяк»
10	Снежинск	1	0,662	0,449	1,816	Всероссийский научно-исследовательский институт технической физики
11	Трехгорный	1	0,006	0,071	3,498	ФГУП «Приборостроительный завод»
12	Бакал	1	0,706	0,337	1,865	ООО «Бакальское рудоуправление»
13	Миныйя	2	0,838	0,437	1,377	ОАО «Миныйярский карьер», ООО «Биыяковский щебеночный завод»
14	Сим	1	0,943	0,89	0,364	ОАО «Агрегат»
	ИТОГО	15				

В базе для Свердловской области изначально рассматривалось 51280 предприятий, которые осуществляли свою деятельность в 17 моногородах. Из первоначальной выборки не рассматривались предприятия, не имевшие выручки в период 2013-2017 гг. Также из выборки был исключен Волчанск. Общая характеристика выборки за 2017 год представлена в таблице 3.

Таблица 3 - Общая характеристика выборки предприятий по моногородам Свердловской области в 2017 г.

№	Моногород	Число МПП	CR	HHI	E	Градообразующее предприятие
1	Асбест	1	0,565	0,319	2,463	ОАО «Уральский асбестовый горнообогатительный комбинат» (ОАО «Ураласбест»)
2	Верхняя Пышма	1	0,598	0,205	2,532	ОАО «Уралэлектромедь»
3	Верхняя Салда	1	0,711	0,547	1,061	АО «Корпорация ВСМПО-АВИСМА»
4	Верхняя Тура	1	0,002	0,327	1,466	ОАО «Верхнетуринский машиностроительный завод»
5	Каменск-Уральский	4	0,926	0,859	0,494	ОАО «Синарский трубный завод (СинТЗ)»; ОАО «Каменск-Уральский завод по обработке цветных металлов»; ОАО «Каменск-Уральский металлургический завод»; ОАО СУАЛ
6	Карпинск	3	0,212	0,07	3,409	ОАО «Электромашиностроительный завод», ООО «Завод горного машиностроения», ООО "Машиностроительный завод «Звезда»
7	Качканар	1	0,813	0,663	1,117	ОАО «ЕВРАЗ Качканарский горнообогатительный комбинат»
8	Краснотурьинск	1	0,251	0,072	4,02	Богословский алюминиевый завод
9	Красноуральск	1	0,926	0,859	0,494	ОАО «Святогор»
10	Малышева	1	0,754	0,572	1,303	ОАО «Малышевское рудоуправление»
11	Нижний Тагил	1	0,2	0,254	3,234	ОАО «Научнопроизводственная корпорация „Уралвагонзавод“ имени Ф. Э. Дзержинского»
12	Первоуральск	1	0,539	0,225	2,87	ОАО «Первоуральский новотрубный завод»
13	Полевской	1	0,67	0,362	2,413	ОАО «Северский трубный завод»
14	Ревда	2	0,845	0,355	1,692	ОАО «Среднеуральский медеплавильный завод», ОАО «Ревдинский завод по обработке цветных металлов»
15	Североуральск	1	0,862	0,745	1,692	ОАО «Севуралбокситруда»
16	Серов	2	0,841	0,354	1,729	ОАО «Металлургический завод имени А. К. Серова», ОАО «Серовский завод ферросплавов»
	ИТОГО	23				

В процессе расчетов показателей выявилась проблема, которая заключается в том, что не для всех моногородов мы можем сделать расчеты. Проблема касалась тех моногородов, градообразующим предприятием которых является филиалом другой организации, расположенной вне муниципального образования. Отчетности филиалов предприятий не публикуются. В данной работе рассматривается один из таких городов г. Тайга, градообразующим предприятием которого является филиал открытого акционерного общества «Российские железные дороги».

Была поставлена задача построить модель, которая могла бы предсказать долю моноотрасли Тайгинского городского округа. Для этого были сделаны следующие шаги:

1. Выгрузка Excel файлов с данными из системы СПАРК.
2. Объединение всех выгруженных файлов в один, оставляя для анализа только необходимые колонки с данными.
3. Расчет выручки моноотрасли, группируя градообразующие предприятия по их ОКПО. Рассчитываем необходимые показатели на основе отобранных данных.
4. Строим матрицу корреляций.
5. На основе полученной матрицы корреляций, отбираем переменные для построения моделей.
6. Строим модели для предсказания доли моноотрасли Тайгинского городского округа.

Вся работа с данными выполнялась на языке программирования Python.

На основе полученных данных из системы СПАРК были рассчитаны следующие показатели для 2013-2017 гг.: выручка, основные средства, оплата труда, открытые предприятия, закрытые предприятия, действующие предприятия, доля моноотрасли (выручка), доля моноотрасли (оплата труда), доля моноотрасли (основные средства), ННІ основные средства, ННІ выручка, ННІ оплата труда, Е основные средства, Е выручка, Е оплата труда, основные средства охранных предприятий, основные средства медицинских

предприятий, основные средства строительных предприятий, основные средства телекоммуникаций, выручка охранных предприятий, выручка медицинских предприятий, выручка строительных предприятий, выручка телекоммуникаций, оплата труда охранных предприятий, оплата труда медицинских предприятий, оплата труда строительных предприятий, оплата труда телекоммуникаций. Количество показателей составило 27.

С рассчитанными показателями была построена матрица корреляций, визуализация матрицы корреляций выполнена в виде тепловой карты, фрагмент матрицы корреляций на рис. 15, полная матрица корреляций в Приложении А.

	Доля моноот расли ОС	Доля моноотр асли Выручка	Доля моноотр асли Оплата труда	НН ОС	НН Выр учка	НН Оплата труда	Е ОС	Е Выручка	Е Оплата труда
Выручка	-0,10	-0,08	-0,08	-0,24	-0,25	-0,34	0,36	0,38	0,53
Основные средства	0,02	0,02	0,00	-0,17	-0,22	-0,34	0,26	0,33	0,53
Оплата труда	-0,02	-0,02	-0,01	-0,20	-0,25	-0,31	0,32	0,38	0,51
Открытые предприятия	-0,19	-0,20	-0,17	-0,30	-0,34	-0,38	0,45	0,51	0,60
Закрытые предприятия	-0,13	-0,15	-0,15	-0,21	-0,28	-0,32	0,35	0,43	0,50
Действующие предприятия	-0,22	-0,22	-0,20	-0,32	-0,37	-0,41	0,49	0,56	0,63
Доля моноотрасли ОС	1,00	0,86	0,73	0,67	0,52	0,37	-0,66	-0,52	-0,34
Доля моноотрасли Выручка	0,86	1,00	0,76	0,61	0,64	0,38	-0,61	-0,63	-0,37
Доля моноотрасли Оплата труда	0,73	0,76	1,00	0,42	0,42	0,40	-0,44	-0,43	-0,35
НН ОС	0,67	0,61	0,42	1,00	0,79	0,73	-0,91	-0,74	-0,69
НН Выручка	0,52	0,64	0,42	0,79	1,00	0,74	-0,77	-0,90	-0,73
НН Оплата труда	0,37	0,38	0,40	0,73	0,74	1,00	-0,70	-0,72	-0,93
Е ОС	-0,66	-0,61	-0,44	-0,91	-0,77	-0,70	1,00	0,87	0,76
Е Выручка	-0,52	-0,63	-0,43	-0,74	-0,90	-0,72	0,87	1,00	0,81
Е Оплата труда	-0,34	-0,37	-0,35	-0,69	-0,73	-0,93	0,76	0,81	1,00
Основные средства охранных предприятий	-0,02	-0,04	-0,05	-0,16	-0,20	-0,29	0,23	0,28	0,40
Основные средства медицинских предприятий	-0,04	-0,09	-0,12	-0,12	-0,20	-0,29	0,25	0,35	0,45
Основные средства строительных предприятий	-0,06	-0,03	-0,03	-0,20	-0,17	-0,22	0,23	0,20	0,29
Основные средства телекоммуникаций	-0,12	-0,07	-0,05	-0,42	-0,38	-0,42	0,49	0,44	0,52
Выручка охранных предприятий	-0,13	-0,13	-0,11	-0,30	-0,31	-0,39	0,40	0,44	0,59
Выручка медицинских предприятий	-0,09	-0,12	-0,12	-0,19	-0,27	-0,33	0,35	0,44	0,53
Выручка строительных предприятий	-0,18	-0,19	-0,18	-0,22	-0,28	-0,32	0,38	0,46	0,54
Выручка телекоммуникаций	-0,04	-0,05	-0,09	-0,18	-0,26	-0,33	0,29	0,36	0,48
Оплата труда охранных предприятий	-0,05	-0,06	-0,01	-0,25	-0,24	-0,29	0,32	0,34	0,46
Оплата труда медицинских предприятий	-0,04	-0,12	-0,13	-0,14	-0,23	-0,30	0,28	0,37	0,49
Оплата труда строительных предприятий	-0,14	-0,14	-0,16	-0,15	-0,21	-0,27	0,31	0,37	0,46
Оплата труда телекоммуникаций	-0,06	0,03	0,05	-0,29	-0,18	-0,23	0,23	0,12	0,23

Рисунок 15 – Фрагмент матрицы корреляций

Исходя из матрицы корреляций, для построения моделей для предсказания доли моноотрасли были выбраны следующие параметры: Действующие предприятия и ННІ Выручка, ННІ Основные средства и ННІ Оплата труда. Так как Тайгинский городской округ не содержит данных о своем комплексе структурных подразделений и, следовательно, данные градообразующего предприятия не участвуют в расчете индекса Херфиндаля-Хиршмана. Чтобы избежать искажений при подстановке данных Тайгинского городского округа в полученную модель было предложено использовать модифицированный индекс Херфиндаля-Хиршмана, т.е. вычесть квадрат доли моноотрасли из выражения:

$$HHI' = \sum_{i=1}^n Y_i^2 - CR^2,$$

где Y_i - доля i -го предприятия в общем объеме моногорода, n – общее число предприятий, CR – доля моноотрасли.

Для проверки качества моделей были рассчитаны коэффициент детерминации и среднеквадратичное отклонение.

Для того чтобы построить модели необходимо разделить данные на обучающую и тестовую выборки (рис. 16).

```
Y = file['Доля моноотрасли ОС'].values
X = file[['ННІ ОС', 'ННІ Выручка', 'Е ОС', 'Е Выручка']].values
```

```
# Split into 70% training and 30% testing set
X_train, X_test, y_train, y_test = train_test_split(X, Y, test_size = 0.3, random_state = 42)
```

Рисунок 16 – Программный код разделения выборки на обучающую и тестовую

Доля основных средств моноотрасли

По выбранным параметрам было получено четыре модели, программный код представлен на рисунке 17.


```

slr = LinearRegression()
slr.fit(X_train, y_train)
MLP = MLPRegressor(hidden_layer_sizes=(4500,), max_iter=5000)
MLP.fit(X_train,y_train)
KNN = KNeighborsRegressor(n_neighbors=2)
KNN.fit(X_train, y_train)
RFR = RandomForestRegressor(n_estimators=150, max_features='auto')
RFR.fit(X_train,y_train)

```

Рисунок 17 – Программный код для построения моделей для доли моноотрасли по основным средствам

Полученная модель линейной регрессии для доли моноотрасли по основным средствам имеет следующий вид:

$$Y(X) = 1,18 - 0,29 \cdot X_1 - 1,15 \cdot X_2,$$

где Y – доля основных средств моноотрасли, X_2 – число действующих предприятий, X_1 – ННГ' Основные средства.

Коэффициент детерминации и среднеквадратичное отклонение полученных моделей представлены в сравнительной таблице 4.

Таблица 4 – Сравнительная таблица полученных моделей для предсказания доли моноотрасли по основным средствам

Название алгоритма машинного обучения	Обучающая выборка		Тестовая выборка	
	R^2	Среднеквадратичное отклонение	R^2	Среднеквадратичное отклонение
Линейная регрессия (метод наименьших квадратов)	0,444	0,061	0,5	0,055
Нейронные сети прямого распространения	0,613	0,043	0,591	0,044
Метод ближайших соседей	0,977	0,003	0,942	0,006
Случайный лес	0,989	0,001	0,952	0,005

Наибольший коэффициент детерминации и наименьшее среднеквадратичное отклонение показал случайный лес с числом деревьев 150, а число признаков для выбора расщепления 150.

Подставив в модель данные трех моногородов, например, Яшкино, Новокузнецк и Киселевск, долю основных средств моноотрасли которых мы знаем, были получены следующие результаты (таблица 5).

Таблица 5 – Предсказанная и рассчитанная доля основных средств моноотрасли Яшкино, Новокузнецка и Киселевска

Моногород		2013	2014	2015	2016	2017
Яшкино	Рассчитанное значение	0,97	0,97	0,96	0,97	0,98
	Предсказанное значение моделью	0,94	0,97	0,97	0,97	0,98
Новокузнецк	Рассчитанное значение	0,35	0,32	0,34	0,31	0,29
	Предсказанное значение моделью	0,35	0,33	0,35	0,32	0,31
Киселевск	Рассчитанное значение	0,86	0,89	0,89	0,88	0,85
	Предсказанное значение моделью	0,88	0,89	0,89	0,85	0,85

Для Тайгинского городского округа были предсказаны следующие значения (таблица 6).

Таблица 6 – Предсказанные моделью доли моноотрасли по основным средствам для Тайгинского городского округа

	2013	2014	2015	2016	2017
Предсказанное значение моделью	0,79	0,78	0,69	0,60	0,57

Доля выручки моноотрасли

Для предсказания доли моноотрасли по выручке было получено четыре модели (рис. 18), результаты представлены в сравнительной таблице 7.

```
slr = LinearRegression()
slr.fit(X_train, y_train)
MLP = MLPRegressor(hidden_layer_sizes=(300,), max_iter=1000)
MLP.fit(X_train, y_train)
KNN = KNeighborsRegressor(n_neighbors=2)
KNN.fit(X_train, y_train)
RFR = RandomForestRegressor(n_estimators=200, max_features='sqrt')
RFR.fit(X_train, y_train)
```

Рисунок 18 - Программный код для построения моделей для доли моноотрасли по выручке

Полученная модель линейной регрессии для доли моноотрасли по основным средствам имеет следующий вид:

$$Y(X) = 1,19 - 0,35 \cdot X_1 - 1,47 \cdot X_2,$$

где Y – доля выручки моноотрасли, X_1 – Действующие предприятия, X_2 – ННГ' Основные средства.

Таблица 7 – Сравнительная таблица полученных моделей для предсказания доли моноотрасли по выручке

Название алгоритма машинного обучения	Обучающая выборка		Тестовая выборка	
	R^2	Среднеквадратичное отклонение	R^2	Среднеквадратичное отклонение
Линейная регрессия (метод наименьших квадратов)	0,427	0,054	0,518	0,051
Нейронные сети прямого распространения	0,471	0,050	0,434	0,050
Метод ближайших соседей	0,962	0,004	0,904	0,006
Случайный лес	0,982	0,002	0,917	0,005

Наибольший коэффициент детерминации и наименьшее среднеквадратичное отклонение показал случайный лес с числом деревьев 200, а число признаков для выбора расщепления $\sqrt{200}$.

Подставив в модель данные трех моногородов, например, Яшкино, Новокузнецк, Киселевск, долю выручки моноотрасли, которых уже были рассчитаны, были получены следующие результаты (таблица 8).

Таблица 8 – Предсказанная и рассчитанная доля выручки моноотрасли для Яшкино, Новокузнецка и Киселевска

Моногород		2013	2014	2015	2016	2017
Яшкино	Рассчитанное значение	0,87	0,95	0,97	0,97	0,96
	Предсказанное значение моделью	0,87	0,94	0,96	0,95	0,88
Новокузнецк	Рассчитанное значение	0,28	0,27	0,26	0,23	0,22
	Предсказанное значение моделью	0,25	0,25	0,24	0,25	0,25
Киселевск	Рассчитанное значение	0,62	0,69	0,70	0,71	0,70
	Предсказанное значение моделью	0,67	0,74	0,73	0,72	0,75

Для Тайгинского городского округа были предсказаны следующие значения (таблица 9).

Таблица 9 - Предсказанные моделью доли моноотрасли по выручке для Тайгинского городского округа

	2013	2014	2015	2016	2017
Предсказанное значение моделью	0,54	0,59	0,46	0,34	0,30

Доля оплаты труда моноотрасли

Для предсказания доли моноотрасли по оплате труда было получено четыре модели (рис. 19), результаты представлены в сравнительной таблице 7.

```
slr = LinearRegression()
slr.fit(X_train, y_train)
MLP = MLPRegressor(hidden_layer_sizes=(300,), max_iter=5000)
MLP.fit(X_train, y_train)
KNN = KNeighborsRegressor(n_neighbors=2)
KNN.fit(X_train, y_train)
RFR = RandomForestRegressor(n_estimators=250, max_features='auto')
RFR.fit(X_train, y_train)
```

Рисунок 19 - Программный код для построения моделей для доли оплаты труда моноотрасли

Полученная модель линейной регрессии для доли моноотрасли по оплате труда имеет следующий вид:

$$Y(X) = -0,43 - 0,43 \cdot X_1 - 1,32 \cdot X_2,$$

где Y – доля оплаты труда моноотрасли, X_1 – число действующих предприятий, X_2 – ННІ Оплата труда, X_2 – ННІ' Выручка.

Таблица 10 – Сравнительная таблица полученных моделей для предсказания доли моноотрасли по оплате труда

Название алгоритма машинного обучения	Обучающая выборка		Тестовая выборка	
	R^2	Среднеквадратичное отклонение	R^2	Среднеквадратичное отклонение
Линейная регрессия (метод наименьших квадратов)	0,527	0,053	0,5	0,045
Нейронные сети	0,049	0,561	0,536	0,041

прямого распространения				
----------------------------	--	--	--	--

Продолжение таблицы 10

Метод ближайших соседей	0,973	0,003	0,871	0,011
Случайный лес	0,984	0,002	0,903	0,009

Наибольший коэффициент детерминации и наименьшее среднеквадратичное отклонение показал случайный лес с числом деревьев 250, а число признаков для выбора расщепления 250.

Подставив в модель данные трех моногородов, Яшкино, Новокузнецк, Киселевск, долю оплаты труда моноотрасли которых уже были нами рассчитаны, были получены следующие значения доли оплаты труда моноотрасли (таблица 11).

Таблица 11 – Предсказанная и рассчитанная доля оплаты труда моноотрасли для Яшкино, Новокузнецка и Киселевска

Моногород		2013	2014	2015	2016	2017
Яшкино	Рассчитанное значение	0,88	0,88	0,91	0,91	0,90
	Предсказанное значение моделью	0,87	0,87	0,90	0,90	0,86
Новокузнецк	Рассчитанное значение	0,37	0,40	0,46	0,44	0,42
	Предсказанное значение моделью	0,39	0,39	0,43	0,43	0,42
Киселевск	Рассчитанное значение	0,70	0,71	0,72	0,78	0,81
	Предсказанное значение моделью	0,72	0,72	0,72	0,78	0,78

Для Тайгинского городского округа были предсказаны следующие значения (таблица 12).

Таблица 12 - Предсказанные моделью доли моноотрасли по оплате труда для Тайгинского городского округа

	2013	2014	2015	2016	2017
Предсказанное значение моделью	0,67	0,72	0,55	0,47	0,46

Во всех трех моделях по предсказанию доли моноотрасли по выручке, основным средствам и оплате труда хорошо показал себя алгоритм случайный лес. Таким образом, используя данный алгоритм, были получены

«скрытые» доли моноотрасли Тайгинского городского округа. Данная модель хорошо себя показала при предсказании доли моноотрасли моногорода, доли моноотрасли которого известны. Максимальная погрешность составила 0,06.

Таким образом в разделе 3 была описана полученная база данных на основе финансовой отчетности предприятий из системы СПАРК, проведен корреляционный анализ из 27 показателей. Были построены модели машинного обучения, используя алгоритмы: линейная регрессия (метод наименьших квадратов), случайный лес, метод ближайших соседей, нейронная сеть прямого распространения. Алгоритм случайный лес показал наилучшие результаты качества построения модели и далее использовался для предсказания доли «скрытой» моноотрасли Тайгинского городского округа.

4 Финансовый менеджмент, ресурсоэффективность и ресурсосбережение

Введение

Целью проекта является на основе обработки данных финансовой отчетности предприятий и применения методов машинного обучения разработать регрессионную модель оценки уровня диверсификации экономики моногорода при отсутствии данных о моноотрасли. Перед тем, как представить продукт на рынке информационных систем, необходимо оценить данную разработку с точки зрения ее востребованности, перспективности, планирования финансовой и коммерческой ценности.

Данный раздел, предусматривает рассмотрение следующих задач:

- Оценка коммерческого потенциала разработки.
- Планирование научно-исследовательской работы;
- Расчет бюджета научно-исследовательской работы;
- Определение ресурсной, финансовой, бюджетной эффективности исследования.

Анализ конкурентных технических решений

Для анализа рассмотрим включение, исключение данных или замена их средними по таким моногородам (если градообразующее предприятие является филиалом или структурным подразделением другой организации, расположенной вне муниципального образования). Анализ представлен в таблице 13.

Таблица 13 - Оценочная карта для сравнения конкурентных технических решений, где K_1 – открытые данные, Ф – разработка НИ

Критерии оценки	Вес критерия	Баллы		Конкурентоспособность	
		БФ	БК1	КФ	КК1
Критерии оценки эффективности					
Повышение качества данных	0,1	5	3	0,5	0,3
Возможность сравнения с другими регионами	0,2	5	2	1	0,4
Точность вычисления	0,2	5	3	1	0,6
Экономические критерии оценки эффективности					

Продолжение таблицы 13

Конкурентоспособность продукта	0,2	5	2	1	0,4
Цена	0,1	5	5	0,5	0,5
Финансирование научной разработки	0,2	3	5	0,6	1
Итого:	1	28	22	4,6	3,2

Таким образом, можно сделать вывод, что наша разработка по многим показателям является более предпочтительной, чем включение, исключение данных или замена их средними (значение 5 является максимальным).

SWOT-анализ

SWOT-анализ применяют для исследования внешней и внутренней среды проекта. SWOT анализ для разработанной регрессионной модели представлен в таблице 14.

Таблица 14 – SWOT-анализ

	Сильные стороны научно-исследовательского проекта: С1. Предсказание доли моноотрасли для моногородов с отсутствием данных о градообразующем предприятии. С2. Единое представление об оценке и анализе диверсификации моногородов. С3.Выявление закономерности развития моногородов.	Слабые стороны научно-исследовательского проекта: Сл1. При большом объеме информации требование мощной вычислительной техники. Сл2. Недостаточность данных по филиалам компаний в моногородах
Возможности: В1. Использование данной разработки муниципалитетами моногородов. В2. Развитие моногородов.	В1С3. Составление более эффективной программы развития моногородов. В2С2. Анализ и корректировка программы развития моногородов.	В1Сл2. Предоставление филиалами отчетности муниципалитетов моногородов.
Угрозы: У1. Отсутствие спроса на новые технологии У2. Несвоевременное финансовое обеспечение научного исследования со стороны государства	У2С3. Обеспечение поддержки областей, в состав которых входят моногорода.	У2Сл1. Привлечение новых источников финансирования.

Планирование научно-исследовательских работ

Структура работ в рамках научного исследования

Комплекс предполагаемых работ включает в себя следующие задачи:

- определить структуру работ в рамках исследования;
- определить участников каждой работы;
- установить продолжительность работ;
- построить график проведения отдельных этапов исследования.

Для выполнения данного исследования (проекта) необходимо сформировать рабочую группу, в состав которой входят научный руководитель (НР) и инженер (И). Для каждой из запланированных работ, необходимо выбрать исполнителя этой работы.

Разработанный список задач и производимых работ, в рамках проектирования, а также распределение исполнителей по этим работам, представлен в виде таблицы 15.

Таблица 15. Комплекс работ по разработке проекта

Основные этапы	№ раб.	Содержание работ	Должность исполнителей
Подготовительный	1	Составление и утверждение задания ВКР	И, НР
	2	Календарное планирование работ по теме	И
	3	Подбор и изучение материалов по теме	И
Исследование и анализ предметной области	4	Анализ исходных данных	И
	5	Выбор метода выполнения работ	И, НР
Теоретические и экспериментальные исследования	6	Написание программы	И
	7	Тестирование программы	И
Обобщение и оценка результатов	8	Анализ полученных результатов работы	И, НР
	9	Составление отчета по работе	И

Определение трудоемкости выполнения работ

Трудовые затраты в большинстве случаев образуют основную часть стоимости разработки, поэтому важным моментом является определение трудоемкости работ каждого из участников научного исследования.

Трудоемкость выполнения научного исследования оценивается экспертным путем в человеко-днях и носит вероятностный характер, т.к. зависит от множества трудно учитываемых факторов. Для определения ожидаемого (среднего) значения трудоемкости $t_{ожі}$ используется следующая формула:

$$t_{ожі} = \frac{3t_{mini} + 2t_{maxi}}{5},$$

где $t_{ожі}$ – ожидаемая трудоемкость выполнения i -ой работы чел.-дн.;

t_{mini} – минимально возможная трудоемкость выполнения заданной i -ой работы (оптимистическая оценка: в предположении наиболее благоприятного стечения обстоятельств), чел.-дн.;

t_{maxi} – максимально возможная трудоемкость выполнения заданной i -ой работы (пессимистическая оценка: в предположении наиболее неблагоприятного стечения обстоятельств), чел.-дн.

Исходя из ожидаемой трудоемкости работ, определяется продолжительность каждой работы в рабочих днях T_p , учитывающая параллельность выполнения работ несколькими исполнителями:

$$T_{p_i} = \frac{t_{ожі}}{Ч_i},$$

где T_{p_i} – продолжительность одной работы, раб. дн.;

$t_{ожі}$ – ожидаемая трудоемкость выполнения одной работы, чел.-дн.;

$Ч_i$ – численность исполнителей, выполняющих одновременно одну и ту же работу на данном этапе, чел.

Разработка диаграммы Ганта

Одним из наиболее удобных и наглядных способов представления календарного плана работы является построение ленточного графика проведения ВКР в форме диаграммы Ганта.

Диаграмма Ганта - это горизонтальный ленточный график, на котором работы по теме представляются протяженными во времени отрезками, характеризующимися датами начала и окончания выполнения данных работ.

Для удобства построения графика, длительность каждого из этапов работ из рабочих дней следует перевести в календарные дни. Для этого необходимо воспользоваться следующей формулой:

$$T_{k_i} = T_{p_i} \cdot k_{\text{кал}},$$

где T_{k_i} – продолжительность выполнения i -й работы в календарных днях;

T_{p_i} – продолжительность выполнения i -й работы в рабочих днях;

$k_{\text{кал}}$ – коэффициент календарности, который определяется по следующей формуле:

$$k_{\text{кал}} = \frac{T_{\text{кал}}}{T_{\text{кал}} - T_{\text{вых}} - T_{\text{пр}}},$$

где $T_{\text{кал}}$ – количество календарных дней в году (365);

$T_{\text{вых}}$ – количество выходных дней в году (52 дня при шестидневной рабочей неделе);

$T_{\text{пр}}$ – количество праздничных дней в году (14).

Таким образом, коэффициент календарности $k_{\text{кал}}$ равен 1,23. Временные показатели проведения научной работы представлены на таблице 16.

Таблица 16. Временные показатели проведения научной работы

№ раб.	Трудоемкость работ, чел-дни			Исполнители	T_{p_i}	T_{k_i}
	$t_{\min i}$	$t_{\max i}$	$t_{\text{ожс}i}$			
1	1	5	3	И, НР	1	1
2	2	3	2	И	2	3
3	20	25	22	И	22	27
4	15	20	17	И	17	21
5	1	3	2	И, НР	1	1
6	15	25	19	И	19	23
7	2	7	4	И	4	5
8	5	7	6	И, НР	3	4
9	7	10	8	И	8	10

Таблица 17. Календарный план-график проведения работ

№ раб.	Наименование работы	Исполнители	T_{k_i} , кал-дн	Продолжительность выполнения работ, дни								
				Март			Апрель		Май			
				1	3	27	21	1	23	5	4	10
1	Составление и утверждение задания ВКР	И, НР	1	■								
2	Календарное планирование работ по теме	И	3		■							
3	Подбор и изучение материалов по теме	И	27			■						
4	Анализ исходных данных	И	21				■					
5	Выбор метода выполнения работ	И, НР	1					■				
6	Написание программы	И	23						■			
7	Тестирование программы	И	5							■		
8	Анализ результатов работы	И, НР	4								■	
9	Составление отчета по работе	И	10									■



- Научный руководитель



- Инженер

Бюджет научно-исследовательского проекта

При планировании бюджета научно-исследовательского проекта должно быть обеспечено полное и достоверное отражение всех видов расходов, связанных с его выполнением. Определение полных затрат на выполнение ВКР производится путем суммирования расходов по следующим статьям:

- материальные затраты;
- основная заработная плата исполнителей;
- дополнительная заработная плата исполнителей;
- отчисления во внебюджетные фонды (страховые отчисления);
- услуги сторонних организаций;
- накладные расходы.

Расчет материальных затрат

Покажем отражение стоимости всех материалов, используемых при разработке проекта, включая расходы на их приобретение и, при необходимости, доставку. Расчет затрат на материалы производится по форме, приведенной в таблице 18.

Таблица 18. Материальные затраты

Наименование	Единица измерения	Количество	Цена за ед., руб.	Сумма, руб.
Бумага	пачка	1	300	300
Картридж для принтера	шт.	1	6000	6000
Канцелярские принадлежности	шт.	1	300	300
Итого				6600

Расчет заработной платы для исполнителей

В данной статье расходов планируется и учитывается основная заработная плата исполнителей, непосредственно участвующих в проектировании выпускной квалификационной работы:

$$Z_{\text{зп}} = Z_{\text{осн}} + Z_{\text{доп}},$$

где $Z_{\text{осн}}$ – основная заработная плата, $Z_{\text{доп}}$ – дополнительная заработная плата.

Основная заработная плата ($Z_{\text{осн}}$) руководителя (лаборанта, инженера) рассчитывается по следующей формуле:

$$Z_{\text{осн}} = Z_{\text{дн}} \cdot T_p,$$

где T_p – продолжительность работ, выполняемых научно-техническим работником, раб. дн.; $Z_{\text{дн}}$ – среднедневная заработная плата работника, руб.

Среднедневная заработная плата рассчитывается по формуле:

$$Z_{\text{дн}} = \frac{Z_m \cdot M}{F_o},$$

где F_o – действительный годовой фонд рабочего времени научно-технического персонала, раб. дн. (таблица 19); M – количество месяцев работы без отпуска в течение года: при отпуске в 48 раб. дней $M=10,4$ месяца, 6-дневная неделя; Z_m – месячный должностной оклад работника, руб.

Таблица 19. Баланс рабочего времени

Показатели рабочего времени	Научный руководитель	Инженер
Календарное число дней	365	365
Количество нерабочих дней: <ul style="list-style-type: none">• выходные дни и праздничные дни	66	66
Потери рабочего времени <ul style="list-style-type: none">• отпуск• невыходы по болезни	52	52
Действительный годовой фонд рабочего времени	248	248

Месячный должностной оклад работника:

$$З_{\text{м}} = З_{\text{тс}} \cdot (1 + k_{\text{пр}}) \cdot k_{\text{р}},$$

где $З_{\text{тс}}$ – заработная плата по тарифной ставке руководителя, руб.;

$k_{\text{пр}}$ – премиальный коэффициент, равный 0,3;

$k_{\text{р}}$ – районный коэффициент, равный 1,3 г.Томск.

Таблица 20. Расчет основной заработной платы

Исполнители	$З_{\text{ок}}$, руб	$k_{\text{пр}}$	$k_{\text{р}}$	$З_{\text{м}}$, руб	$З_{\text{дн}}$, руб	$T_{\text{р}}$, дн	$З_{\text{осн}}$, руб
Руководитель	33300	0,3	1,3	56277	2360	5	11800
Инженер	14874	0,3	1,3	25137,06	1054,13	77	81168,01

Дополнительная заработная плата

Дополнительная заработная плата определяется по формуле:

$$З_{\text{доп}} = k_{\text{доп}} \cdot З_{\text{осн}},$$

где $k_{\text{доп}}$ - коэффициент дополнительной заработной платы ($k_{\text{доп}} = 0,15$).

Дополнительная заработная плата руководителя:

$$З_{\text{доп}} = 0,15 \cdot 11800 = 1770 \text{ рублей.}$$

Дополнительная заработная плата инженера:

$$З_{\text{доп}} = 0,15 \cdot 81168,01 = 12175,2 \text{ рублей.}$$

Отчисления во внебюджетные фонды

Отчисления во внебюджетные фонды являются обязательными по установленным законодательством Российской Федерации нормам органам государственного социального страхования (ФСС), пенсионного фонда (ПФ) и медицинского страхования (ФФОМС) от затрат на оплату труда работников.

Величина отчислений во внебюджетные фонды определяется исходя из следующей формулы:

$$З_{\text{внеб}} = k_{\text{внеб}} \cdot (З_{\text{осн}} + З_{\text{доп}}),$$

Где $k_{внеб}$ – коэффициент отчислений на уплату во внебюджетные фонды (пенсионный фонд, фонд обязательного медицинского страхования и пр.). В соответствии с Федеральным законом от 24.07.2009 №212-ФЗ установлен размер страховых взносов равный 30%.

Отчисления во внебюджетные фонды представлены в таблице 21.

Таблица 21. Отчисления во внебюджетные фонды

Исполнитель	Заработная плата ($З_{осн} + З_{доп}$), руб.
Научный руководитель	13570
Инженер	93343,21
Коэффициент отчислений во внебюджетные фонды	0,3
Отчисления во внебюджетные фонды	
Научный руководитель	4071
Инженер	28002,96
Итого отчислений во внебюджетные фонды	32073,96

Услуги сторонних организаций и накладные расходы

Накладные расходы учитывают прочие затраты организации, не попавшие в предыдущие статьи расходов: печать и ксерокопирование материалов исследования, оплата услуг связи, электроэнергии, почтовые и телеграфные расходы, размножение материалов и т.д. Рассчитаем затраты на электроэнергию, потребляемую оборудованием.

Компьютер потребляет примерно 220 Вт/ч, учитывая 6 часов в день непрерывной работы с компьютером получаем $220 \cdot 6 \cdot 77 = 10164$ Вт, $101,64 \cdot 5,8 = 589,51$ руб.

Величину коэффициента накладных расходов возьмем в размере 16%.

$$З_{накл} = (\text{сумма статей } 1 \div 7) \cdot k_{нр}$$

$$З_{накл} = (6600 + 13570 + 93343,21 + 32073,96 + 589,51) \cdot 0,16 = 23388,27 \text{ руб.}$$

Формирование бюджета затрат НИ

Рассчитанная величина затрат научно-исследовательской работы является основой для формирования бюджета затрат проекта, который при формировании договора с заказчиком защищается научной организацией в качестве нижнего предела затрат на разработку научно-технической продукции. Определение бюджета затрат на научно-исследовательский проект приведен в таблице 22.

Таблица 22. Расчет бюджета затрат НИ

Наименование статьи	Сумма, руб.
Материальные затраты	6600
Основная заработная плата	92968,01
Дополнительная заработная плата	13945,2
Отчисления во внебюджетные фонды	32073,96
Накладные расходы	23388,27
Бюджет затрат НИ	168975,44

Определение ресурсной (ресурсосберегающей), финансовой, бюджетной, социальной и экономической эффективности исследования

Определение эффективности происходит на основе расчета интегрального показателя эффективности научного исследования. Его нахождение связано с определением двух средневзвешенных величин: финансовой эффективности и ресурсоэффективности.

Интегральный показатель финансовой эффективности научного исследования получают в ходе оценки бюджета затрат трех (или более) вариантов исполнения научного исследования. Для этого наибольший интегральный показатель реализации технической задачи принимается за базу расчета (как знаменатель), с которым соотносятся финансовые значения по всем вариантам исполнения.

Интегральный финансовый показатель разработки определяется как:

$$I_{\text{финр}}^{\text{исп.}i} = \frac{\Phi_{pi}}{\Phi_{\text{max}}},$$

где $I_{\text{финр}}^{\text{исп.}i}$ – интегральный финансовый показатель разработки; Φ_{pi} – стоимость i -го варианта исполнения; Φ_{max} – максимальная стоимость исполнения научно-исследовательского проекта (в т.ч. аналоги). За максимально возможную стоимость исполнения примем 200000 руб.

Интегральный показатель ресурсоэффективности вариантов исполнения объекта исследования можно определить следующим образом:

$$I_{pi} = \sum a_i \cdot b_i,$$

где I_{pi} – интегральный показатель ресурсоэффективности для i -го варианта исполнения разработки; a_i – весовой коэффициент i -го варианта исполнения разработки; b_i^a, b_i^p – бальная оценка i -го варианта исполнения разработки, устанавливается экспертным путем по выбранной шкале оценивания; n – число параметров сравнения.

Расчет интегрального показателя ресурсоэффективности представлен в таблице 23.

Таблица 23. Расчет интегрального показателя ресурсоэффективности.

Критерии	Весовой коэффициент параметра	Оценка	Максимальная оценка
Количество рассчитанных показателей экономической концентрации	0,2	5	5
Требуется наличия большого количества данных	0,25	5	5
Простота применения	0,15	4	5
Конкурентоспособность (с другими ресурсами)	0,25	4	5
Возможность применения любыми муниципалитетами	0,15	3	5
ИТОГО:	1	4,3	5

$$I_{\text{ршис}} = 0,2 \cdot 5 + 0,25 \cdot 5 + 0,15 \cdot 4 + 0,25 \cdot 4 + 0,15 \cdot 3 = 4,3;$$

$$I_{\text{ршис}} = 0,2 \cdot 5 + 0,25 \cdot 5 + 0,15 \cdot 5 + 0,25 \cdot 5 + 0,15 \cdot 5 = 5$$

Интегральный показатель эффективности вариантов исполнения разработки ($I_{исп.}$) определяется на основании интегрального показателя ресурсоэффективности и интегрального финансового показателя по формуле:

$$I_{исп.1} = \frac{I_{p-исп1}}{I_{финр.1}}, \quad I_{исп.2} = \frac{I_{p-исп2}}{I_{финр.2}} \text{ и т.д.}$$

Сравнение интегрального показателя эффективности вариантов исполнения разработки позволит определить сравнительную эффективность проекта и выбрать наиболее целесообразный вариант из предложенных. Сравнительная эффективность проекта ($\Theta_{ср}$):

$$\Theta_{ср} = \frac{I_{исп.1}}{I_{исп.2}}.$$

Так как исследование выполнено в одном варианте исполнения, рассчитаем интегральный показатель эффективности относительно максимально возможного варианта. Сравнительная эффективность разработки представлена в таблице 24.

Таблица 24. Сравнительная эффективность разработки

Показатели	Исп.1	Исп. max
Интегральный финансовый показатель	0,84	1
Интегральный показатель ресурсоэффективности разработки	4,3	5
Интегральный показатель эффективности	5	5
Сравнительный показатель эффективности	1	

Сравнение значений интегральных показателей эффективности позволяет понять и выбрать более эффективный вариант решения поставленной в диссертации технической задачи с позиции финансовой и ресурсной эффективности.

Заключение по разделу

В результате выполнения данного раздела были получены следующие результаты:

1. Проведен анализ конкурентных технических решений, в результате которого получили, что наша разработка по многим показателям является более предпочтительной.
2. Проведен SWOT-анализ, были определены сильные и слабые стороны, возможности и угрозы.
3. При проведении планирования был разработан план-график проведения работ для научного руководителя и инженера. Были определены общее количество календарных дней для выполнения работы – 77 дней.
4. Сформирован бюджет затрат НИ, которые составили 168975,44 рублей.
5. Определены интегральный финансовый показатель, интегральный показатель эффективности, интегральный показатель эффективных вариантов исполнения. Рассчитанные показатели показали, что НИ эффективно с позиции финансовой и ресурсной эффективности.

5 Социальная ответственность

Введение

Целью проекта является на основе обработки данных финансовой отчетности предприятий и применения методов машинного обучения разработать регрессионную модель оценки уровня диверсификации экономики моногорода при отсутствии данных о моноотрасли.

Работа выполнялась на компьютере, который состоит из системного блока и монитора, работа производится сидя, при небольшом физическом напряжении. Рабочее место представляет собой компьютерный стол с персональным компьютером в лаборатории НТБ Томского политехнического университета.

Данный раздел посвящен анализу вредных и опасных факторов производственной среды для операторов ПЭВМ, а также разработке программ по минимизации воздействия вредоносного и опасного влияния выявленных факторов, снижению вредных воздействий на окружающую среду и защите в чрезвычайных ситуациях.

Правовые и организационные вопросы обеспечения безопасности

Регулирование отношений между работником и работодателем, касающихся оплаты труда, трудового распорядка, особенности регулирования труда женщин, детей, людей с ограниченными способностями и проч., осуществляется законодательством РФ, а именно трудовым кодексом РФ [43]. Продолжительность рабочего дня не должна быть меньше указанного времени в договоре, но не больше 40 часов в неделю. Для работников до 16 лет – не более 24 часов в неделю, от 16 до 18 лет и инвалидов I и II группы – не более 35 часов. Возможно установление неполного рабочего дня для беременной женщины; одного из родителей (опекуна, попечителя), имеющего ребенка в возрасте до четырнадцати лет

(ребенка–инвалида в возрасте до восемнадцати лет). Оплата труда при этом производится пропорционально отработанному времени, без ограничений оплачиваемого отпуска, исчисления трудового стажа и других прав. В течение рабочего дня работнику должен быть предоставлен перерыв для отдыха и питания продолжительностью не более двух часов и не менее 30 минут, который в рабочее время не включается. Всем работникам предоставляются выходные дни, работа в выходные дни осуществляется только с письменного согласия работника. Выполняемая работа по виду трудовой деятельности относится к группе Б – работа по вводу информации и по категории тяжести к категории В – суммарное время непосредственной работы с ПЭВМ за рабочую смену не более 6 часов за смену. Требования к организации рабочих мест пользователей:

- Рабочее место должно быть организовано с учетом требований согласно ГОСТ 12.2.032–78 «ССБТ. Рабочее место при выполнении работ сидя. Общие эргономические требования» и ГОСТ 12.2.061–81 «ССБТ. Оборудование производственное. Общие требования безопасности к рабочим местам» [44,45].
- Конструкция рабочей мебели (рабочий стол, кресло, подставка для ног) должна обеспечивать возможность индивидуальной регулировки соответственно росту пользователя и создавать удобную позу для работы. Вокруг ПК должно быть обеспечено свободное пространство не менее 60–120 см.
- На уровне экрана должен быть установлен оригинал–держатель. На рисунке 1 схематично представлены требования к рабочему месту.



Рисунок 20 – Организация рабочего места

Работа программиста связана с постоянной работой за компьютером, следовательно, могут возникать проблемы, связанные со зрением. Также неправильная рабочая поза может оказывать негативное влияние на здоровье. Таким образом, неправильная организация рабочего места может послужить причиной нарушения здоровья и появлением психологических расстройств. Согласно СанПиН 2.2.2/2.4.1340–03 «Гигиенические требования к персональным электронно-вычислительным машинам и организации работы» [46]:

- яркость дисплея не должна быть слишком низкой или слишком высокой;
- размеры монитора и символов на дисплее должны быть оптимальными;
- цветовые параметры должны быть отрегулированы таким образом, чтобы не возникало утомления глаз и головной боли;
- опоры для рук не должны мешать работе на клавиатуре;
- верхний край монитора должен находиться на одном уровне с глазом, нижний – примерно на 20° ниже уровня глаза;
- дисплей должен находиться на расстоянии 45–60 см от глаз;
- локтевой сустав при работе с клавиатурой нужно держать под углом 90° ;

- монитор должен иметь антибликовое покрытие;
- работа за компьютером не должна длиться более 6 часов, при этом необходимо каждые 2 часа делать перерывы по 15–20 минут;
- высота стола и рабочего кресла должны быть комфортными.

Производственная безопасность

Для обеспечения производственной безопасности необходимо проанализировать воздействия на человека вредных и опасных производственных факторов, которые могут возникать при разработке или эксплуатации проекта «Разработка информационной системы управления эффективностью диверсификации экономики моногорода». Производственный фактор считается вредным, если воздействие этого фактора на работника может привести к его заболеванию. Возможные опасные и вредные факторы согласно ГОСТ 12.0.003–2015 [47] приведены в таблице 25.

Таблица 25 – Возможные опасные и вредные факторы

Факторы (ГОСТ 12.0.003–2015)	Этапы работ			Нормативные документы
	Разработка	Внедрение	Эксплуатация	
1.Отклонение показателей микроклимата	+	+	+	СанПиН 2.2.4.548–96 [47]
2.Превышение уровня шума	+	+	+	ГОСТ 12.1.003–2014 [48] СанПиН 2.2.4.3359–16 [49]
3.Недостаточность освещенность зоны	+	+	+	СанПиН 2.2.1/2.1.1.1278–03 [50]
4.Повышенный уровень электромагнитных полей	+	+	+	ГОСТ 12.1.006–84 [51] СанПиН 2.2.4.3359–16 [49]
5. Повышенное значение напряжения в электрической цепи, замыкание которой может произойти через тело человека.	+	+	+	СанПиН 2.2.2/2.4.1340–03 [52]

Далее более подробно рассмотрены опасные и вредные факторы и мероприятия по снижению их воздействия.

Отклонение показателей микроклимата

Микроклимат производственных помещений – это климат внутренней среды этих помещений, который определяется действующими на организм человека сочетаниями температуры, влажности и скорости движения воздуха, а также интенсивности теплового излучения от нагретых поверхностей. Оптимальные значения этих характеристик зависят от сезона (холодный, тёплый), а также от категории физической тяжести работы. Согласно требованиям, СанПиН 2.2.4.548–96. «Гигиенические требования к микроклимату производственных помещений», оптимальные и допустимые параметры микроклимата в офисах приведены в таблицах 26 и 27.

Таблица 26 – Оптимальные значения характеристик микроклимата

Период года	Кат. раб.	Температура воздуха, °С	Относительная влажность воздуха, %	Скорость движения воздуха, м/с
Холодный	Ia	22-24	40-60	0,1
Теплый	Ia	23-25	40-60	0,1

Таблица 27 – Допустимые величины показателей микроклимата

Период года	Кат. раб.	Температура воздуха, °С	Относительная влажность воздуха, %	Скорость движения воздуха, м/с
Холодный	Ia	19-24	15-75	0,1-0,2
Теплый	Ia	20-28	15-75	0,1-0,3

Если температура воздуха отличается от нормальной, то время пребывания в таком помещении должно быть ограничено в зависимости от категории тяжести работ.

Для создания благоприятных условий труда и повышения производительности, необходимо поддерживать оптимальные параметры микроклимата рабочего помещения, в котором происходит разработка информационной системы управления эффективностью диверсификации экономики моногорода. Для этого предусмотрены следующие средства:

центральное отопление, вентиляция (искусственная и естественная), искусственное кондиционирование. Все из вышеперечисленных средств используются на рабочем месте.

Можно отметить, что фактические значения температуры, влажности и скорости движения воздуха соответствуют нормам, установленным в СанПиН 2.2.4.548–96. Для соответствия этим нормам производится проветривание помещения каждый час и ежедневная влажная уборка.

Превышение уровня шума

Шум определяется как звук, оцениваемый негативно и наносящий вред здоровью, например, ухудшение слуха, негативное воздействие на центральную и вегетативную нервную системы. Основными источниками шума в помещении являются:

- система охлаждения центральных процессоров;
- жесткие диски.

При выполнении основной работы на ПЭВМ по разработке информационной системы управления эффективностью диверсификации экономики моногорода, уровень шума на рабочем месте не должен превышать 50 дБА. Допустимые уровни звукового давления в помещениях для персонала, осуществляющего эксплуатацию ЭВМ при разных значениях частот согласно СН 2.2.4/2.1.8.562–96 «Шум на рабочих местах, в помещениях жилых, общественных зданий и на территории жилой застройки», приведены в таблице 28.

Таблица 28 – Допустимые уровни звука на рабочем месте

Вид трудовой деятельности, рабочее место	Уровни звукового давления, дБ, в октавных полосах со среднегеометрическими частотами, Гц									Уровни звука и эквивалентного звука, дБА
	31,5	63	125	250	500	1000	2000	4000	8000	
Научная деятельность, проектирование, программирование	86	71	61	54	49	45	42	40	38	50

Для снижения уровня шума, производимого персональными компьютерами, регулярно проводилось техническое обслуживание:

- чистка от пыли;
- замена смазывающих веществ;
- применение звукопоглощающих материалов.

Недостаточность освещенности рабочей зоны

Недостаточная освещенность рабочей зоны является вредным производственным фактором, возникающим при работе с ПЭВМ, уровни которого регламентируются СП 52.13330.2016. Причиной являются недостаточность естественного освещения, искусственного освещения, пониженная контрастность. Недостаточность освещения снижает производительность труда, увеличивает утомляемость, количество допускаемых ошибок, ухудшению зрения.

Установлено, что свет, помимо обеспечения зрительного восприятия, воздействует на нервную оптико–вегетативную систему, систему формирования иммунной защиты, рост и развитие организма и влияет на многие основные процессы жизнедеятельности, регулируя обмен веществ и устойчивость к воздействию неблагоприятных факторов окружающей среды. Освещение при разработке информационной системы управления эффективностью диверсификации экономики моногорода должно включать в себя как естественное, так и искусственное. Для источников искусственного освещения применяют люминесцентные лампы типа ЛБ [48].

Минимальный размер объекта различия входит в диапазон 0,5 до 1,0, следовательно, работа относится к разряду IV. Подразряд Г, т.к. контраст объектов различия с фоном большой, сам фон светлый. В соответствии с СП 52.13330.2016 уровень искусственного освещения должен быть не менее 300 лк. Пульсация при работе с компьютером не должна превышать 5%.

Увеличение коэффициента пульсации освещенности снижает зрительную работоспособность, повышает утомляемость, воздействует на нервные элементы коры головного мозга и фоторецепторные элементы сетчатки глаз. Для снижения пульсации необходимо использовать светильники, в которых лампы работают от переменного тока частотой 300 Гц и выше, так как при такой частоте пульсация не оказывает влияния на общую и зрительную работоспособность.

Повышенный уровень электромагнитных полей

Воздействие электромагнитного излучения на человека зависит от напряженностей электрического и магнитного полей, потока энергии, частоты колебаний, размера облучаемого тела. Работа по разработке информационной системы управления эффективностью диверсификации экономики моногорода проводилась на современном компьютере, где значения электромагнитного излучения малы и отвечают требованиям СанПиН 2.2.1/2.1.1.1278–12. «Электромагнитные поля в производственных условиях», которые приведены в таблице 29.

Таблица 29 – Временные допустимые уровни электромагнитных полей

Наименование параметров		Временные допустимые уровни электромагнитных полей
Напряженность электрического поля	в диапазоне частот 5 Гц–2 кГц	25 В/м
	в диапазоне частот 2 кГц–400 кГц	2,5 В/м
Плотность магнитного потока	в диапазоне частот 5 Гц–2 кГц	250 нТл
	в диапазоне частот 2 кГц–400 кГц	25 нТл
Поверхность видеомонитора	электростатический потенциал экрана	500 В

Основной способ снижения вредного воздействия (ухудшение зрения) – это увеличение расстояния от источника (не менее 50 см от пользователя). При работе за компьютером специальные экраны и другие средства индивидуальной защиты применены не были.

Повышенное значение напряжения в электрической цепи, замыкание которой может произойти через тело человека

Поражение электрическим током является опасным производственным фактором, поскольку оператор ПЭВМ имеет дело с электрооборудованием в процессе разработки информационной системы управления эффективностью диверсификации экономики моногорода. Нормы электробезопасности на рабочем месте регламентируются СанПиН 2.2.2/2.4.1340–03, вопросы требований к защите от поражения электрическим током освещены в ГОСТ 12.1.019–2017 ССБТ [49].

Помещение, где расположено рабочее место разработчика проекта, относится к помещениям без повышенной опасности ввиду отсутствия следующих факторов: сырость, токопроводящая пыль, токопроводящие полы, высокая температура, возможность одновременного прикосновения человека к имеющим соединение с землей металлоконструкциям зданий, технологическим аппаратам, механизмам и металлическим корпусам электрооборудования.

Для разработчика проекта при работе с электрическим оборудованием обязательны следующие меры предосторожности: перед началом работы нужно убедиться, что выключатели и розетка закреплены и не имеют оголённых токоведущих частей, при обнаружении неисправности оборудования и приборов необходимо, вызвать мастера ответственного за оборудование.

Экологическая безопасность

В данном разделе рассматривается воздействие на окружающую среду деятельности по разработке проекта «Разработка информационной системы управления эффективностью диверсификации экономики моногорода», а также самого продукта информационной системы в результате его реализации.

При написании ВКР были определены следующие источники загрязнения окружающей среды: бумажные черновики, картриджи, лампочки. Все три вида должны быть утилизированы. Бумажные отходы содержат конфиденциальную информацию, недопустимую для распространения другим лицам. Шредер является машиной по измельчению бумаги. Полученные после размельчения отходы подлежат сдаче в макулатуру для дальнейшей их утилизации специальными средствами.

Картридж необходимо разобрать на составляющие - фотобарабан, вал первичного заряда, лезвие очистки, уплотнительное лезвие барабана, магнитный вал, лезвие дозировки тонера, уплотнительная чека. Все детали для утилизации сортируются в соответствие с техническими характеристиками (материал, из которого они изготовлены). Вышедшие из использования лампочки подлежат обязательной сдаче в приемный пункт завода светотехники [54].

Безопасность в чрезвычайных ситуациях

Наиболее вероятной чрезвычайной ситуацией является возникновение пожара, так как на рабочем месте в лаборатории НТБ Томского политехнического университета располагается большое количество ЭВМ. В соответствии с нормами пожарной безопасности, помещения с ЭВМ относятся к категории В (пожароопасные).

Основные причины возникновения возгораний: нарушение правил эксплуатации электрического оборудования, эксплуатация его в неисправном состоянии, перегрузка электрических сетей, применение неисправных осветительных приборов, электропроводки и устройств. Для предупреждения возгораний в помещении необходимо соблюдать: установленный режим эксплуатации электрических устройств, противопожарные нормы и правила при установке оборудования, проводить технические осмотры и планово-предупредительные ремонты оборудования и технических средств

противопожарной защиты и пожаротушения (огнетушители) согласно утвержденному графику. В помещении должен быть установлен углекислотный огнетушитель типа ОУ–5 для тушения пожаров и расположен план пожарной эвакуации на видных местах. В случае угрозы возникновения ЧС необходимо отключить электропитание, вызвать по телефону пожарную команду, эвакуировать людей из помещения согласно плану эвакуации. При наличии небольшого очага пламени можно воспользоваться подручными средствами с целью прекращения доступа воздуха к объекту возгорания.

Выводы по разделу: в ходе работы были рассмотрены и выявлены опасные и вредные факторы на рабочем месте разработчика в процессе работы над проектом «Разработка информационной системы управления эффективностью диверсификации экономики моногорода», которые могут стать причиной профессиональных заболеваний и травм. Были разработаны меры предосторожности и профилактические работы по устранению угроз для здоровья человека. Рассмотрены правовые и организационные вопросы обеспечения безопасности, а также безопасность в чрезвычайных ситуациях. В результате анализа предлагаемого программного продукта данной ВКР, вредных факторов, влияющих на окружающую среду не обнаружено.

Заключение

В Российской Федерации существует большое количество моногородов, особенно сконцентрированы они в Кемеровской, Челябинской и Свердловских областях.

В ходе работы были рассмотрены и рассчитаны показатели концентрации и диверсификации на основе данных: о выручке, основных средствах и оплате труда предприятий моногородов. Для расчета данных показателей была разработана программа для их быстрой обработки и расчетов. Для моногородов, градообразующее предприятие которых является филиалом другого предприятия, расположенной вне муниципального образования, невозможно рассчитать данные показатели. К данным моногородам относится Тайгинский городской округ.

На основе рассчитанных данных была построена матрица корреляций по полученным показателям. Проанализировав матрицу корреляций, были выбраны наиболее коррелированные показатели для последующего анализа. Для Тайгинского городского округа была предложена модель предсказания доли «скрытой» моноотрасли. Из рассмотренных алгоритмов машинного обучения наиболее эффективным показал себя алгоритм случайный лес. Используя данную модель были восстановлены нулевые значения доли моноотрасли по выручке, основным средствам и оплате труда. Что повысило качество данных о Тайгинском городском округе. Повысив качество данных по Тайгинскому городскому округу, у моногорода появляется возможность сопоставить себя с другими моногородами и регионами.

Полученные результаты показывают, что цель была достигнута, а поставленные задачи реализованы.

Список публикаций студента

1. Антонова И. С., Малеева Е. А. Предприятия-флагманы в моногородах [Электронный ресурс] // Информационные технологии в науке, управлении, социальной сфере и медицине: сборник научных трудов V Международной конференции: в 2 т., Томск, 17-21 Декабря 2018. - Томск: ТПУ, 2018 - Т. 2 - С. 9-11.
2. Антонова И. С., Малеева Е. А. Эффективность предприятий-флагманов на основе портфельной теории // Информационные технологии в науке, управлении, социальной сфере и медицине: сборник научных трудов VI Международной научной конференции, Томск, 14-19 Октября 2019. - Томск: ТПУ, 2019 - С. 199-201.
3. I. Antonova, E. Pchelintsev, E. Maleeva Internal market concentration of company towns // XIV International Conference “Russian Regions in the Focus of Changes”. - Ekaterinburg, Russian Federation, 14-16 November, 2019 – P. 48-53.
4. I. Antonova, E. Maleeva IT services and the digital transformation of the regions with high concentration of single-industry towns towards smart city // 2019 International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON). - 2019 – P. 50-54.

Список литературы

1. Рощина И.В., Дятлова Н.А. Влияние градообразующего предприятия на социально-трудовую сферу монопрофильного муниципального образования // Известия Иркутской государственной экономической академии. – 2013. - №3. – С. 25-29.
2. Ряховская А.Н., Полякова А.Г. Модернизация российских моногородов: стратегические ориентиры институционального реформирования // Эффективное антикризисное управление. – 2016. - №4(97). – С. 54-65.
3. Землянский Д.Ю., Ламанов С.В. Сценарии развития монопрофильных городов России // Вестн. Моск. ун-та. Сер. 5. География. 2014. №4. С. 69-74.
4. Dale B. An institutional approach to local restructuring: the case of four Norwegian mining towns. European Urban and Regional Studies, 9, no. 1 (2002): 5–20.
5. Mingaleva Z. A., Sheresheva M. Y., Oborin M. S., Gvarliani T. E. Networking of small cities to gain sustainability. Entrepreneurship and Sustainability Issues, 5, no. 1 (2017): 140–156.
6. Оборин М.С., Пахалов А.М., Шерешева М.Ю. Эффективность стратегического планирования развития малых городов на основе сетевого механизма координации // Вестн. Моск. ун-та, Сер.6. Экономика. 2017. №4. С. 100-117.
7. Christiaensen L., Todo Y. Poverty reduction during the rural–urban transformation—the role of the missing middle // World Development. 2014. № 63. P. 43–58.
8. Ferré C., Ferreira F., Lanjouw P. Is There a Metropolitan Bias? The Relationship between Poverty and City Size in a Selection of Developing Countries // World Bank Economic Review. 2012. Vol. 26. № 3. P. 351–382.

9. Данилов-Данильян В.И. Экономико-математический энциклопедический словарь. М.: Инфра-М, 2003. 668 с.
10. Дондоков Б.Б. Формирование туристическо-рекреационного кластера как путь диверсификации региональной экономики // Проблемы современной экономики, 2010. - № 3 (35).
11. Ускова Т.В., Копасова С.С. Диверсификация экономики региона важнейшее условие повышения ее конкурентоспособности // Экономика региона: проблемы и перспективы развития, 2007. - №40.
12. Innis H.A. Settlement and the mining frontier // Settlement and the Forest Frontier in Eastern Canada / Ed. by A.R.M. Lower. – Toronto: Macmillan of Canada, 1936. – P. 170–407.
13. Landis P.H. Three Iron Mining Towns: A Study in Cultural Change. – Edwards Brothers, Inc., 1938. – 210 p.
14. Гранберг А.Г. Артоболевская С., Ковалёва Г. Реструктуризация старопромышленных регионов: опыт России и мира // Региональное развитие и сотрудничество. 1998. № 1-2. С. 4-23.
15. Правительство Российской Федерации. Постановление от 29 июля 2014 г. №709 «О критериях отнесения муниципальных образований Российской Федерации к монопрофильным (моногородам) и категориях монопрофильных муниципальных образований Российской Федерации (моногородов) в зависимости от рисков ухудшения их социально-экономического положения».
16. Allen J.B. The Company Town in the American West. – University of Oklahoma Press, 1966. – 120 p.
17. Bebbington A., Hinojosa L., Bebbington D.H. et al. Contention and ambiguity: Mining and the possibilities of development // Development and Change. – 2008. – No. 39 (6). – P. 887–914.
18. Tonts M., Plummer P., Lawrie, M. Socio-economic wellbeing in Australian mining towns: A comparative analysis // Journal of Rural Studies. – 2012. – No. 28 (3). – P. 288–301.

19. Alpaydin E. Introduction to machine learning 2nd edition. // The MIT Press, 2010. – P. 1-13.
20. Breiman, L. Random Forests. Machine Learning 45, 5–32 (2001).
21. Hastie, T., Tibshirani R., Friedman J. Chapter 15. Random Forests // The Elements of Statistical Learning: Data Mining, Inference, and Prediction. — 2nd ed. — Springer-Verlag, 2009. — 746 p.
22. Altman, Naomi S. (1992). "An introduction to kernel and nearest-neighbor nonparametric regression" (PDF). The American Statistician. 46 (3): 175–185.
23. Jaskowiak, Pablo A.; Campello, Ricardo J. G. B. "Comparing Correlation Coefficients as Dissimilarity Measures for Cancer Classification in Gene Expression Data". Brazilian Symposium on Bioinformatics (BSB 2011): 1–8.
24. Coomans, Danny; Massart, Desire L. (1982). "Alternative k-nearest neighbour rules in supervised pattern recognition: Part 1. k-Nearest neighbour classification by using alternative voting rules". Analytica Chimica Acta. 136: 15–27.
25. Нейронная сеть // Большая российская энциклопедия : [в 35 т.] / гл. ред. Ю. С. Осипов. — М.: Большая российская энциклопедия, 2004-2017.
26. Мак-Каллок У. С., Питтс В. Логическое исчисление идей, относящихся к нервной активности на Wayback Machine // Автоматы / Под ред. К. Э. Шеннона и Дж. Маккарти. — М.: Изд-во иностр. лит., 1956. — С. 363—384.
27. Агентство «Росбизнесконсалт» Москва, 1995 [Электронный ресурс]. - URL: <https://www.rbc.ru/newspaper/2019/06/24/5d0cbafc9a7947af87e8c419> (дата обращения: 30.05.2020).
28. Хаггет П. Пространственный анализ в экономической географии. — Москва: Прогресс, 1968

- 29.Об основах федеральной поддержки депрессивных территорий российской федерации : Проект федерального закона 2001 г. № 91010-3. – Режим доступа: <http://consultant.ru>
- 30.Сошкин И.А. Стратегическое планирование устойчивого развития муниципального образования: дисс. канд. экон. наук. – Тула, 2003. – 128 с.
- 31.Большой экономический словарь / Под ред. А.Н. Азриляна. – 7-е зд., доп. – Москва : Институт новой экономики, 2007. – 1472 с.
- 32.Ансофф И. Стратегическое управление: Сокр. пер. с англ. / Науч. ред. и авт. предисл. Л.И. Евенко. – Москва: Экономика, 1989. – 519 с.
- 33.Томпсон, А.А., Стрикленд А.Дж. Стратегический менеджмент: Искусство разработки и реализации стратегии. – Москва: ЮНИТИ Банки и биржи, 1998. – 576 с.
- 34.Привалихин В.И. Исторический очерк: Тайга - 100 лет/В.И.Привалихин. - Новосибирск: Фирма «Издатель», 1996. - 375 с.
- 35.Распоряжение Правительства РФ от 29 июля 2014 года № 1398-р «Об утверждении перечня моногородов».
- 36.Merriam-Webster Dictionary. – Access: <http://www.merriamwebster.com/dictionary/company%20town>.
- 37.Green H. The Company Town: The Industrial Edens and Satanic Mills That Shaped the American Economy. – New York: Basic Books, 2010. – P. 7–27.
- 38.Монопрофильные города и градообразующие предприятия // Союз инвесторов. – Режим доступа: http://www.unioninvest.ru/city_mong.html
- 39.Ожегов С.И. Словарь русского языка: 70 000 слов / под ред. Н.Ю. Шведовой. – Москва: Русский язык, 1989. – 924 с.
40. Толковый словарь русского языка: В 4 т. / Под ред. Д. Н. Ушакова. – Москва, 2000.
- 41.Финансовый энциклопедический словарь. Электронная библиотека. – Режим доступа: http://www.bookorbita.com/library/spravochnik_enciklopedii/neizvesten_5/g.html (дата обращения 20.05.2020).

42. Кузнецова, Г.Ю. Социально-экономические трансформации монопрофильных поселений в переходной экономике // Региональные исследования. Смоленский гуманитарный университет. – Смоленск: 2004. – №1(3). – С. 33–43.
43. Трудовой кодекс Российской Федерации от 30.12.2001 N 197–ФЗ (ред. от 16.12.2019)
44. ГОСТ Р ИСО 9241–2–2009. Эргономические требования к проведению офисных работ с использованием видеодисплейных терминалов (VDT).
45. ГОСТ 12.0.003–2015 «Опасные и вредные производственные факторы. Классификация».
46. СанПиН 2.2.4.548–96 Гигиенические требования к микроклимату производственных помещений.
47. ГОСТ 12.1.005–88. Межгосударственный стандарт. ССБТ. Общие санитарно–гигиенические требования к воздуху рабочей зоны (ред. от 20.06.2000) – М.: Изд–во стандартов, 2000. – 75 с.].
48. ГОСТ 12.1.045–84 Система стандартов безопасности труда (ССБТ). Электростатические поля.
49. ГОСТ 12.4.124–83 ССБТ. Средства защиты от статического электричества. Общие технические требования.
50. ГОСТ 12.2.032–78 ССБТ. Рабочее место при выполнении работ сидя. Общие эргономические требования.
51. ГОСТ Р ИСО 9241–2–2009. Эргономические требования к проведению офисных работ с использованием видеодисплейных терминалов (VDT).
52. СП 52.13330.2016 Естественное и искусственное освещение. Актуализированная редакция СНиП 23–05–95.
53. СанПиН 2.1.7.1322–03 «Гигиенические требования к размещению и обезвреживанию отходов производства и потребления»
54. Постановление Правительства Российской Федерации от 3 сентября 2010 г. N 681 г. Москва «Об утверждении Правил обращения с отходами производства и потребления в части осветительных

устройств, электрических ламп, ненадлежащие сбор, накопление, использование, обезвреживание, транспортирование и размещение которых может повлечь причинение вреда жизни, здоровью граждан, вреда животным, растениям и окружающей среде».

Приложение А

Матрица корреляций

	Выручка	Основные средства	Оплата труда	Открытые предприятия	Закрытые предприятия	Действующие предприятия	Доля моноотрасли ОС	Доля моноотрасли Выручка	Доля моноотрасли Оплата труда	ННИ ОС	ННИ Выручка	ННИ Оплата труда	Е ОС	Е Выручка	Е Оплата труда	Основные средства охранных предприятий	Основные средства медицинских предприятий	Основные средства строительных предприятий	Основные средства телекоммуникаций	Выручка охранных предприятий	Выручка медицинских предприятий	Выручка строительных предприятий	Выручка телекоммуникаций	Оплата труда охранных предприятий	Оплата труда медицинских предприятий	Оплата труда строительных предприятий	Оплата труда телекоммуникаций
Выручка	1,00	0,93	0,95	0,89	0,84	0,91	-0,10	-0,08	-0,08	-0,24	-0,25	-0,34	0,36	0,38	0,53	0,70	0,77	0,52	0,55	0,95	0,89	0,91	0,75	0,87	0,81	0,91	0,33
Основные средства	0,93	1,00	0,95	0,86	0,77	0,87	0,02	0,02	0,00	-0,17	-0,22	-0,34	0,26	0,33	0,53	0,72	0,78	0,51	0,61	0,91	0,85	0,86	0,81	0,83	0,79	0,87	0,36
Оплата труда	0,95	0,95	1,00	0,85	0,85	0,90	-0,02	-0,02	-0,01	-0,20	-0,25	-0,31	0,32	0,38	0,51	0,78	0,85	0,47	0,58	0,91	0,92	0,89	0,85	0,80	0,86	0,92	0,26
Открытые предприятия	0,89	0,86	0,85	1,00	0,80	0,97	-0,19	-0,20	-0,17	-0,30	-0,34	-0,38	0,45	0,51	0,60	0,67	0,76	0,43	0,48	0,94	0,90	0,94	0,71	0,82	0,81	0,91	0,20
Закрытые предприятия	0,84	0,77	0,85	0,80	1,00	0,83	-0,13	-0,15	-0,15	-0,21	-0,28	-0,32	0,35	0,43	0,50	0,68	0,78	0,34	0,41	0,83	0,86	0,85	0,70	0,69	0,79	0,85	0,10
Действующие предприятия	0,91	0,87	0,90	0,97	0,83	1,00	-0,22	-0,22	-0,20	-0,32	-0,37	-0,41	0,49	0,56	0,63	0,72	0,82	0,43	0,51	0,93	0,94	0,95	0,77	0,76	0,85	0,93	0,17
Доля моноотрасли ОС	-0,10	0,02	-0,02	-0,19	-0,13	-0,22	1,00	0,86	0,73	0,67	0,52	0,37	-0,66	-0,52	-0,34	-0,02	-0,04	-0,06	-0,12	-0,13	-0,09	-0,18	-0,04	-0,05	-0,04	-0,14	-0,06
Доля моноотрасли Выручка	-0,08	0,02	-0,02	-0,20	-0,15	-0,22	0,86	1,00	0,76	0,61	0,64	0,38	-0,61	-0,63	-0,37	-0,04	-0,09	-0,03	-0,07	-0,13	-0,12	-0,19	-0,05	-0,06	-0,12	-0,14	0,03
Доля моноотрасли Оплата труда	-0,08	0,00	-0,01	-0,17	-0,15	-0,20	0,73	0,76	1,00	0,42	0,42	0,40	-0,44	-0,43	-0,35	-0,05	-0,12	-0,03	-0,05	-0,11	-0,12	-0,18	-0,09	-0,01	-0,13	-0,16	0,05
ННИ ОС	-0,24	-0,17	-0,20	-0,30	-0,21	-0,32	0,67	0,61	0,42	1,00	0,79	0,73	-0,91	-0,74	-0,69	-0,16	-0,12	-0,20	-0,42	-0,30	-0,19	-0,22	-0,18	-0,25	-0,14	-0,15	-0,29
ННИ Выручка	-0,25	-0,22	-0,25	-0,34	-0,28	-0,37	0,52	0,64	0,42	-0,79	1,00	0,74	-0,77	-0,90	-0,73	-0,20	-0,20	-0,17	-0,38	-0,31	-0,27	-0,28	-0,26	-0,24	-0,23	-0,21	-0,18
ННИ Оплата труда	-0,34	-0,34	-0,31	-0,38	-0,32	-0,41	0,37	0,38	0,40	0,73	0,74	1,00	-0,70	-0,72	-0,93	-0,29	-0,29	-0,22	-0,42	-0,39	-0,33	-0,32	-0,33	-0,29	-0,30	-0,27	-0,23
Е ОС	0,36	0,26	0,32	0,45	0,35	0,49	-0,66	-0,61	-0,44	-0,91	-0,77	-0,70	1,00	0,87	0,76	0,23	0,25	0,23	0,49	0,40	0,35	0,38	0,29	0,32	0,28	0,31	0,23
Е Выручка	0,38	0,33	0,38	0,51	0,43	0,56	-0,52	-0,63	-0,43	-0,74	-0,90	-0,72	0,87	1,00	0,81	0,28	0,35	0,20	0,44	0,44	0,44	0,46	0,36	0,34	0,37	0,37	0,12
Е Оплата труда	0,53	0,53	0,51	0,60	0,50	0,63	-0,34	-0,37	-0,35	-0,69	-0,73	-0,93	0,76	0,81	1,00	0,40	0,45	0,29	0,52	0,59	0,53	0,54	0,48	0,46	0,49	0,46	0,23
Основные средства охранных предприятий	0,70	0,72	0,78	0,67	0,68	0,72	-0,02	-0,04	-0,05	-0,16	-0,20	-0,29	0,23	0,28	0,40	1,00	0,89	0,44	0,44	0,69	0,82	0,67	0,85	0,54	0,81	0,79	0,14
Основные средства медицинских предприятий	0,77	0,78	0,85	0,76	0,78	0,82	-0,04	-0,09	-0,12	-0,12	-0,20	-0,29	0,25	0,35	0,45	0,89	1,00	0,34	0,40	0,77	0,93	0,81	0,88	0,58	0,95	0,91	0,00
Основные средства строительных предприятий	0,52	0,51	0,47	0,43	0,34	0,43	-0,06	-0,03	-0,03	-0,20	-0,17	-0,22	0,23	0,20	0,29	0,44	0,34	1,00	0,71	0,38	0,32	0,34	0,61	0,43	0,30	0,40	0,81
Основные средства	0,55	0,61	0,58	0,48	0,41	0,51	-0,12	-0,07	-0,05	-0,42	-0,38	-0,42	0,49	0,44	0,52	0,44	0,40	0,71	1,00	0,48	0,41	0,40	0,69	0,52	0,38	0,43	0,79
Выручка охранных предприятий	0,95	0,91	0,91	0,94	0,83	0,93	-0,13	-0,13	-0,11	-0,30	-0,31	-0,39	0,40	0,44	0,59	0,69	0,77	0,38	0,48	1,00	0,91	0,94	0,69	0,90	0,84	0,91	0,21
Выручка медицинских предприятий	0,89	0,85	0,92	0,90	0,86	0,94	-0,09	-0,12	-0,12	-0,19	-0,27	-0,33	0,35	0,44	0,53	0,82	0,93	0,32	0,41	0,91	1,00	0,94	0,80	0,73	0,95	0,97	0,03
Выручка строительных предприятий	0,91	0,86	0,89	0,94	0,85	0,95	-0,18	-0,19	-0,18	-0,22	-0,28	-0,32	0,38	0,46	0,54	0,67	0,81	0,34	0,40	0,94	0,94	1,00	0,69	0,76	0,85	0,95	0,09
Выручка телекоммуникаций	0,75	0,81	0,85	0,71	0,70	0,77	-0,04	-0,05	-0,09	-0,18	-0,26	-0,33	0,29	0,36	0,48	0,85	0,88	0,61	0,69	0,69	0,80	0,69	1,00	0,56	0,79	0,81	0,35
Оплата труда охранных предприятий	0,87	0,83	0,80	0,82	0,69	0,76	-0,05	-0,06	-0,01	-0,25	-0,24	-0,29	0,32	0,34	0,46	0,54	0,58	0,43	0,52	0,90	0,73	0,76	0,56	1,00	0,68	0,75	0,37
Оплата труда медицинских предприятий	0,81	0,79	0,86	0,81	0,79	0,85	-0,04	-0,12	-0,13	-0,14	-0,23	-0,30	0,28	0,37	0,49	0,81	0,95	0,30	0,38	0,84	0,95	0,85	0,79	0,68	1,00	0,90	0,00
Оплата труда строительных предприятий	0,91	0,87	0,92	0,91	0,85	0,93	-0,14	-0,14	-0,16	-0,15	-0,21	-0,27	0,31	0,37	0,46	0,79	0,91	0,40	0,43	0,91	0,97	0,95	0,81	0,75	0,90	1,00	0,10
Оплата труда телекоммуникаций	0,33	0,36	0,26	0,20	0,10	0,17	-0,06	0,03	0,05	-0,29	-0,18	-0,23	0,23	0,12	0,23	0,14	0,00	0,81	0,79	0,21	0,03	0,09	0,35	0,37	0,00	0,10	1,00

Приложение Б

Modeling of the process of assessing the level of diversification of a single-industry town using the example of Taiginsky district

Студент:

Группа	ФИО	Подпись	Дата
ЗВМ81	Малеева Е. А.		

Консультант ШИП (руководитель ВКР)

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент	Антонова И.С.	к.э.н., доцент		

Консультант – лингвист ШБИП ОИЯ

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Старший преподаватель	Квашнина О.С.			

2 Modeling of the process of assessing the level of diversification of a single-industry town using the example of Taiginsky district

2.1 General characteristics of Taiginsky district

The objectives of this paragraph include the analysis of the socio-economic state of a single-industry town, the analysis of factors affecting its development.

Taiginsky district is a major railway station on the Trans-Siberian railway, a railway junction with a branch line to Tomsk. The foundation for the appearance of Taiga was the laying of the Great Siberian steam railway. The first buildings on the site of the city appeared in 1895. Officially, Taiga was recognized as a settlement of the Tomsk province in 1896 [34]. In 1911, Taiga was converted from the village into an urban settlement, the population of which was 10,100 people. Taiginsky district was established on December 17, 2004.

The 1920 census gave a clearer picture of the Taiga inhabitants. The population was 12,035 people, railway workers with families made up 73% of the total population. Besides railway enterprises, there were also 4 blacksmiths, 2 shoemakers, 2 carpenters and tin workshops, a brick and brewing factory, a watch workshop – a total of 21 enterprises where only 127 people were employed. [36]. Thus, we can conclude that the city existed for the maintenance and operation of a railway station. According to the census of 1937, the population was already 31,600 people. Figure 5 shows the change in population size from 1920 to 2019.

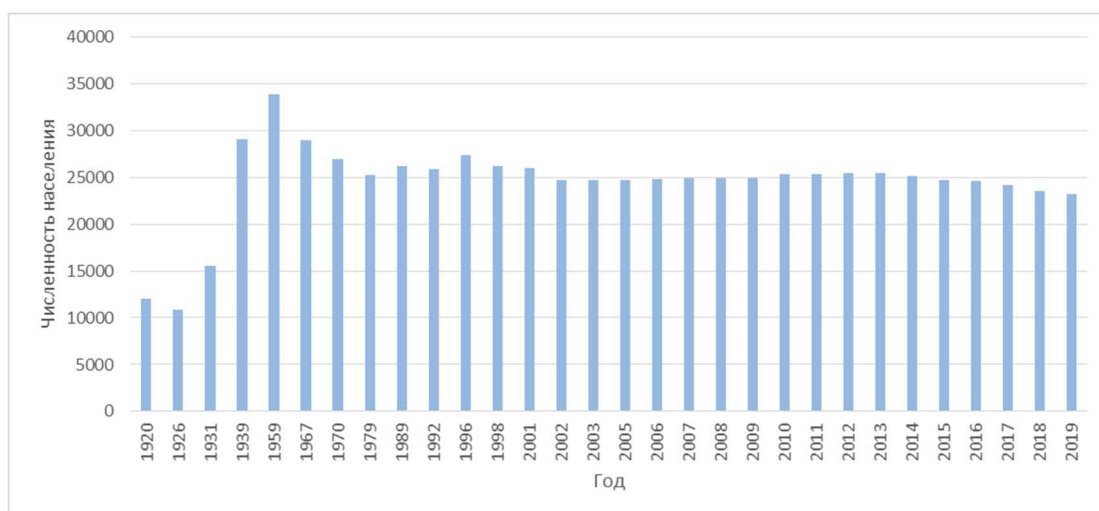


Figure 5 - Dynamics of population

Until 1959, the population grew noticeably. The post-war five-year plans were remarkable for the city for its rapid development and the work of railway workers. New residential areas and industrial enterprises appeared.

In the late 1960s, one after another, new workshops of the depot came into operation: preventive, automatic, electric. The electric traction brought with it telemechanics and automation, radio communication. This made it possible not only to facilitate the work of drivers and improve traffic safety, but also to increase the volume of traffic by 74.3%, reduce the cost of production by 62%.

In 1994, the city industry decreased its production volumes. The city, its economic state, largely depended on the profits on the railway. Freight traffic fell by two-thirds and passenger traffic by 40%. During the crisis years of perestroika, many large enterprises ceased their activities: a plant for the production of mineral wool, a plant for the production of expanded clay, a pasta factory and an asphalt plant.

The restructuring of railway transport, which began in 1997, had a negative impact on the city's economy. All enterprises were stripped of their legal entity status, and in the period from 1998 to 2015, the number of employees at the main enterprise decreased by almost 1,800 people (31.8% of the maximum level in 2006).

The decree of the Russian Federation Government from July 29, 2014 No. 1398-R "On approval of the list of single-industry towns" urban region included in the category "Single-industry municipalities of the Russian Federation (single-industry towns) with risks of social and economic deterioration». The population as of January 1, 2019 was 24,948 people. The city-forming enterprise is "Russian Railways", whose average number of employees is 3,291 people.

Historically, Taiginsky district is connected with railway activity, now its city-forming enterprise is "Russian Railways" and it is included in the list of mono-profile municipalities of the Russian Federation. A variance analysis was conducted for three indicators of enterprises in terms of revenue, fixed assets and wages, a high range of data was observed.

2.2 Systematization of machine learning methods to support decision-making on the level and directions of diversification of single-industry towns

The objectives of the current section include the description of algorithms of machine learning methods that were used to model Taiginsky district share of a single industry.

Regression

Regression is one of the most well-known and understandable algorithms in statistics and machine learning.

For example, we want to have a system that can predict the price of a used car. Input data is the car's attributes: manufacturer, year of manufacture, engine power, mileage, and other information that, in our opinion, affects the cost of the car. The output is the price of the car. Problems where the output number is a regression problem. Let X denote the attributes of the car, and Y denote the price of the car. Again, by investigating past transactions, we can collect training data, and a machine learning program adapts a function to this data to learn Y as a function of X [19]. For example, where the adapted function has the form:

$$y = w_0 + w \cdot x.$$

Both regression and classification are controlled learning tasks, where input is X , output is Y , and the task is to learn the mapping from input to output. The approach in machine learning is that we assume a model defined with precision to a set of parameters:

$$y = g(x | \theta),$$

where $g(\cdot)$ is the model, and θ is its parameters. Y is a number in the regression and the class code (for example, 0/1) in the case of classification. $g()$ is a regression function, or in classification, a discriminant function that separates instances of different classes. The machine learning program optimizes the parameters θ in such a way that the approximation error is minimized, that is, our estimates are as close as possible to the correct values set in the training set. Figure 12 shows the types of linear and nonlinear regression.

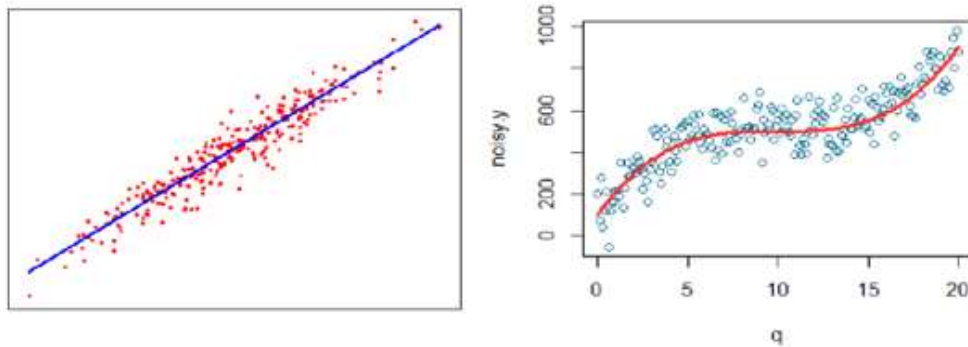


Figure 12 – Types of regression

In cases where the linear model is too restrictive, you can use, for example, a quadratic function:

$$y = w_0 + w_1x + w_2x^2$$

is higher-order polynomial, or any other non-linear function of the input signal, this time optimizing its parameters for the best possible match.

Random forest algorithm

This machine learning algorithm, which was proposed by Leo Breiman and Adele Cutler, uses an ensemble of decision trees [20]. The algorithm includes two ideas: Breiman bagging method and the random subspace method. This algorithm is used for classification, regression, and clustering tasks.

Decision trees are a good family of basic classifiers for bagging. The random subspace method reduces the correlation between trees and allows you to avoid overtraining. The algorithm for constructing a random forest for N trees is as follows [21]:

- 4) For each $n = 1, \dots, N$: generate a sample using X_n a bootstrap;
- 5) Build a decision tree b_n from a sample X_n :
 - the best attribute is selected based on the specified criteria and we divide it in the tree and so on until the selection is exhausted;
 - the tree is built until there are no more n_{\min} objects in each leaf, or until the tree reaches a certain height;
 - for each partition, m random features are selected from the original ones, among which the optimal division is searched. It is recommended to use this method for regression problems $m = \frac{n}{3}$.
- 6) The final classifier: $a(x) = \frac{1}{N} \sum_{i=1}^N b_i(x)$.

Nearest neighbor method

The K-nearest neighbor algorithm (k-NN) is a nonparametric method used for classification and regression [22]. In both cases, the input data consists of k nearest training examples in the feature space. The result depends on whether k-NN is used for classification or regression.

In the k-NN classification, the conclusion is that it belongs to a class. An object is classified by the set of votes of its neighbors, and the object is assigned

to the class most common among its k nearest neighbors (k is a positive integer, usually a small one). If $k = 1$, then the object is simply assigned to the class of this single nearest neighbor.

In the k -NN regression, the output value is the property value for the object. This value is the average value of the values of k nearest neighbors.

The nearest neighbor method is a type of instance-based learning or lazy learning, where the function is only approximated locally, and all calculations are deferred until the function is evaluated.

For both classification and regression, assigning weights to neighbors' contributions can be a useful method, so that closer neighbors contribute more to the average value than more distant ones. For example, the general weighing scheme is to give each neighbor a weight of $1/d$, where d is the distance to the neighbor.

Neighbors are taken from a set of objects for which the class is known (for K -NN classification) or the property value of the object (for K -NN regression).

This can be considered as a training set for the algorithm, although no explicit training step is required.

A special feature of the k -NN algorithm is that it is sensitive to the local data structure.

The method algorithm looks like this:

- Training examples are vectors in a multidimensional space of objects, each of which has a class label. The training phase of the algorithm consists only of storing feature vectors and class labels for training samples.
- At the classification stage, k is a user-defined constant, and an unmarked vector (query or test point) is classified by assigning a label that is most commonly found among the K training samples closest to this query point. A commonly used distance metric for continuous variables is the Euclidean distance. For discrete variables, such as text classification, you can use another metric, such as the overlap metric (or Hamming distance). For

example, in the context of gene expression microchip data, k-NN with correlation coefficients such as Pearson and Spearman [23]. Often, the accuracy of the k-NN classification can be significantly improved if the distance metric is studied using specialized algorithms, such as analyzing the nearest neighbors with a large margin or analyzing neighborhood components.

The disadvantage of the basic classification of "majority vote" occurs when the distribution of classes is distorted. That is, examples of a more frequent class tend to dominate the prediction of a new example, because they tend to be distributed among k nearest neighbors due to their large number [24]. One way to overcome this problem is to weigh the classification based on the distance from the test point to each of its k nearest neighbors. The class (or value, in regression problems) of each of the k nearest points is multiplied by the weight proportional to the inverse distance from this point to the test point. Another way to overcome the skew is to use abstraction in the data representation. For example, in a self-organizing map, each node is a representative (center) of a cluster of similar points, regardless of their density in the source data, k-NN data will be applied to the self-organizing map.

Direct propagation neural network

A neural network is a mathematical model, its software or hardware implementation, built on the principle of organization and functioning of biological neural networks [25]. The first attempt to model the processes taking place in the brain was the neural networks of the human brain by Machalova and U. Pitts [26]. Having developed training algorithms, models were used in forecasting tasks, for pattern recognition, in control tasks, etc.

In a direct propagation neural network, all connections are directed strictly from input neurons to output neurons (Figure 7). Neural networks are often described as a layer cake, where each layer consists of input, hidden, or output

cells. Cells of one layer are not connected to each other, and neighboring layers are usually completely connected.

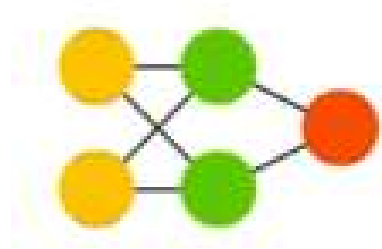


Figure 7 - Diagram of a direct propagation neural network (yellow is the input neuron, green is the hidden neuron, and red is the output neuron)

The simplest neural network has two input cells and one output cell, and can be used as a model of logic gates. A forward propagation neural network is usually trained using the error propagation method of reverse propagation, in which the network receives multiple inputs and outputs. This process is called learning with a teacher, and it differs from learning without a teacher in that in the second case, the network composes a set of output data independently. The above error is the difference between input and output. If the network has a sufficient number of hidden neurons, it is theoretically able to simulate the interaction between input and output data.

Thus, the following machine learning methods were considered: regression, random forest, nearest neighbor method, and direct propagation neural networks.

2.3 Justification of the programming language for developing a machine learning model

Python was chosen as the programming language. Python is the most popular high-level programming language with dynamic semantics.

Python is often used in machine learning since this language is best suited for such tasks as collecting, organizing and analyzing data, and then using the information obtained to create algorithms for artificial intelligence. In addition, it has excellent data processing performance.

One of the main reasons why Python is used for machine learning is that it has many frameworks that simplify the code-writing process and reduce development time. Numpy is used in scientific calculations, SciPy is used in advanced calculations, and SciKit — Learn is used in data extraction and analysis. These libraries work in frameworks such as TensorFlow, CNTK, and Apache Spark.

The following advantages of the Python programming language can be highlighted:

Understandability. The Python programming language is the highest-level language. Simple syntax allows you to test complex algorithms with minimal time spent on their implementation.

Extensive support. A lot of useful resources and high-quality documentation for programmers.

Thus, these factors show why the Python programming language is actively used for machine learning.